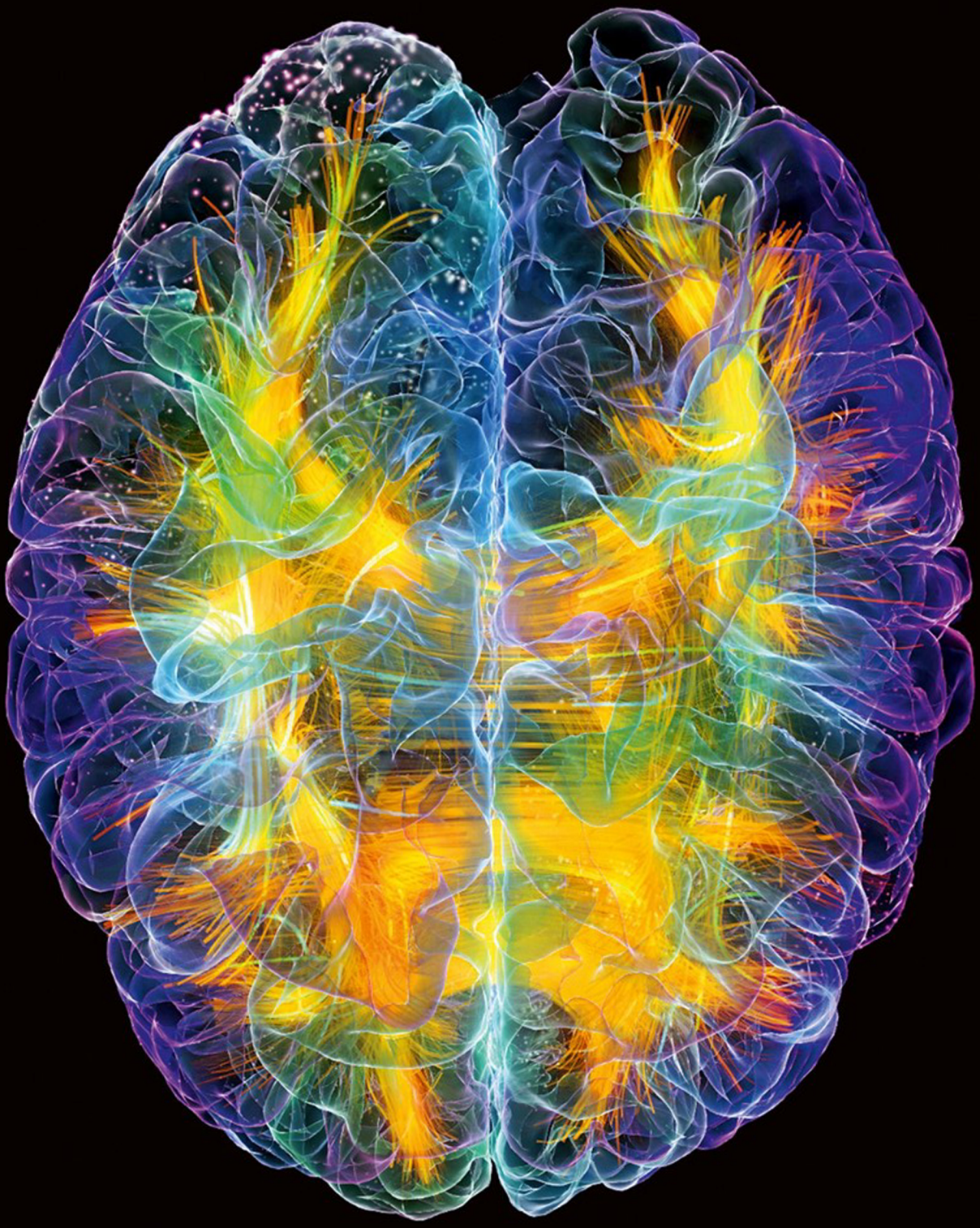
TESIS DOCTORAL

Changes in psychological and biological signals after completing an adaptive training program requiring working memory related cognitive processes

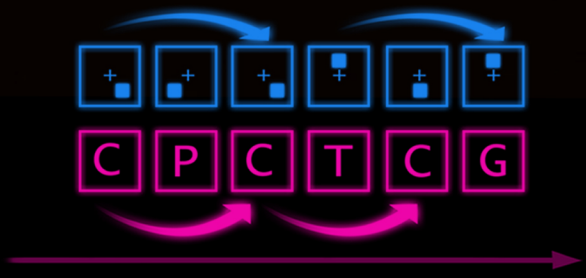
Francisco Javier
Román González



2015

DIRECTOR:
Prof. Roberto Colom

CO-DIRECTORES:
Dr. Sherif Karama
Dra. Kenia Martínez



UAM
UNIVERSIDAD AUTONOMA
DE MADRID

Psicologìa Biologica
y de la Salud

Changes in psychological and biological signals after completing an adaptive training program requiring working memory related cognitive processes.

Author
FRANCISCO JAVIER ROMÁN GONZÁLEZ

Thesis submitted to the Department of Biological and Health Psychology to obtaining the degree of Doctor in Psychology.

Under the supervision of:
Prof. ROBERTO COLOM MARAÑÓN.
Dr. KENIA MARTÍNEZ RODRÍGUEZ.
Dr. SHERIF KARAMA.

November 2015

Agradecimientos/Acknowledgements

Tengo mucho que agradecer a muchas personas.

- Por su trabajo, tiempo y comprensión, al Dr. Roberto Colom. Rob, sin tu ayuda y supervisión esta tesis no habría sido posible. Muchas gracias por crear un ambiente de trabajo estupendo en el que me he sentido 100% realizado. Siempre voy a recordar con mucho cariño nuestras sesudas reuniones dónde nos emocionábamos e inquietábamos con cada resultado obtenido o idea surgida. Porque la verdadera ciencia hay que vivirla y tú la vives como nadie. Gracias por dejarme ser un "*Phileas Fogg*". Esta tesis doctoral es el fin de una etapa, pero no un punto y final. ¡Hablamos pronto del próximo "*paper*"!

- Dra. Martínez, no sabes todo lo que te agradezco. Gracias por ser mi contrapunto, por todas esas tardes en la UAM en las que nos hemos quedado trabajando, por todo ese tiempo que nos hemos dedicado mutuamente. Nunca podré olvidar la primera vez que empezamos a colaborar (¡todavía tengo pesadillas con los surcos!). Kenia, gracias por poner otro punto de vista a esta tesis doctoral.

- Dr. Karama, Thanks! You know that I love Montreal. It is the city of my dreams. Sherif, thanks for your hospitality and for your wise advice reviewing this PhD report! I know our collaboration will improve with time (like our participants with the n-back program).

- Mi amor por la ciencia no comenzó en la UAM. Mi más profundo agradecimiento a dos grandes investigadores que me inculcaron la pasión por la investigación rigurosa y de calidad: el Dr. Luis Aguado y la Dra. Mª Ángeles Quiroga. Luis, gracias por ser quien primero confió en mi y me abrió las puertas de su laboratorio en la UCM. En esas cabinas me forje como investigador. Mª Ángeles, esta tesis ha sido posible gracias a que me abriste un nuevo mundo en tercero de carrera. Gracias por tus sabios consejos y por tu dedicación. Nuestra colaboración nunca acabará porque tienes ganas de volver donde te sientes como en casa.

- Por supuesto, mil gracias al equipo de trabajo que ha colaborado en el desarrollo de esta tesis: Dr. Sergio Escorial, mi maestro y mi alumno a partes iguales. ¡Gracias por todas esas reuniones/charlas improvisadas de camino a la UAM en tu coche!. Dr.

Francisco J. Abad, gracias por ayudarme en los momentos de crisis (no sólo) metodológica; siempre tienes una solución. Dr. Jesús Privado, nuestro n-back no existiría sin ti. Dra. Pei-Chun Shih, gracias por tu colaboración y por permitirme participar en otros excitantes proyectos. Gracias a toda la gente de la Fundación Reina Sofía por su actitud colaboradora. Y por último, pero no menos importante, gracias Dr. Miguel Burgaleta. Mis primeros pasos como neurocientífico fueron bajo tu supervisión, y, por tanto, mis primeras respuestas las obtuve exprimiendo tus neuronas.

- All my gratitude to labmates from the Montreal Neurological Institute (Montreal), the Laboratory of NeuroImaging (Los Angeles) and the Beckman Institute (Urbana-Champaign). Special thanks to Dr. Yasser Iturria-Medina and Dr. Aron Barbey.

- ¡Que sería de mí sin el laboratorio 6! Quiero agradecer a la UAM y al departamento de Psicología Biológica y de la Salud el espacio (no sólo físico) de trabajo dónde he pasado tantas horas. Del lab. 6 no sólo me llevo compañeros de laboratorio, sino también grandes amigos. Gracias a la antigua cúpula: Elena, Edu y Carlos; y a la nueva cúpula: Carlos, Marta, Miriam y Sandra. (Carlos eres un abusón, siempre en la cúpula). Niñxs (jaja) gracias, gracias y mil gracias. Nuestras risas en la cafetería de la UAM (y fuera de la facultad) son inolvidables.

- Y, en el terreno estrictamente personal, gracias a mis compañeros de carrera que me apoyaron en el mundo de la investigación: Isa, Irene, Nuria, Santi, Feli y Helena. A mi equipo/amigos de la UCM por seguir contando conmigo: Bere, Vero, María, Tere y José (María, mil gracias por tu excelente revisión). A mis amigos, Norber, Unai, José, David, Eva, Artur, Rocío, Betty, Elie, Sara, Leire, Andrés, Ana, Irina, Marina, Raúl, Mario, Diego, Murilo, Alex, Dani, Majo, Ángel y un largo etcétera (¡la lista puede ser infinita!) por las veces que me habéis preguntado: "¿Cómo vas?". A Francesco Riga por la portada de la tesis.

- Y, por supuesto, gracias a mi familia por permitirme estar donde estoy y ser quien soy.

Madrid, 20 de Octubre de 2015

General Index

Content Index

Part I: Theory

Chapter 1: Intelligence and Cognitive Processes

Chapter 2: Attempts to raise intelligence

Chapter 3: Neuroimaging Framework for Studying Brain Plasticity

Chapter 4: Framework and Research Hypothesis

Part II: Research

Chapter 5: Methodology and Basic Design

Chapter 6: Training performance (Q1)

Chapter 7: Gray matter responsiveness and n-back training (Q2.1)

Chapter 8: Structural connectivity and n-back training (Q2.2)

Chapter 9: Follow-up (Q3 & Q4)

Chapter 10: General Discussion

Abbreviations

A: Agreeableness.

AC: Anterior commissure.

ACD: Anatomical Connection Density.

ACP: Anatomical Connection Probability.

ACS: Anatomical Connection Strength.

ADD: Attention Deficit Disorder.

ADHD: Attention Deficit Hyperactivity Disorder.

ANCOVA: Analysis of Covariance.

ANOVA: Analysis of Variance.

APA: American Psychological Association.

ATT: Attention Control.

B_0: Magnetic field.

B1: Radiofrequency pulse.

BA: Brodmann area.

C: Conscientiousness.

Cl: Clustering coefficient

CE: Compatibility Effect.

CFA: Confirmatory Factor Analysis.

CFI: Comparative Fit Index.

CGMV: Cortical gray matter volume.

CHC: Cattel-Horn-Carroll model.

CMIN: Chi Square.

CT: Cortical Thickness.

CSA: Cortical Surface Area.

CSF: Cerebrospinal Fluid.

DAT-AR: Differential Aptitude Test: Abstract reasoning scale.

DAT-NR: Differential Aptitude Test: Numerical reasoning scale.

DAT-VR: Differential Aptitude Test: Verbal reasoning scale.

DICOM: Digital Imaging and Communication in Medicine.

DSI: Diffusion spectrum imaging.

DTI: Diffusion tensor model.

DWI: Diffusion weighted images.

E: Extraversion.

EP: Executive Processing.

FA: Fractional anisotropy.

FDR: False Discovery Rate.

FDT: FMRIB's Diffusion Toolbox.

FWE: Family-Wise Error.

g: General Factor of Intelligence.

Gc: Verbal-Crystallized Intelligence.

Gf: Abstract-Fluid Intelligence.

GFA: Generalized Fractional Anisotropy.

GM: Gray Matter.

Gv: Visual-Spatial Intelligence.

HARDI: High-Angular-Resolution Diffusion Imaging.

IRT: Item Response Theory.

IQ: Intelligence Quotient.

L: Path length index.

LTM: Long-Term Memory.

MD: Mean diffusivity.

MLM: Maximum-Likelihood.

MNI: Montreal Neurological Institute.

MRI: Magnetic Resonance Imaging.

N: Neuroticism.

NBS: Network-based Statistic.

NEX: Number of Excitations.

NIFTI: Neuroimaging Informatics Technology Initiative.

NLSY: National Longitudinal Study on Youth.

O: Openness to Experience.

ODF: Orientational Distribution Function.

PC: Posterior Commissure.

P-FIT: Parieto-Frontal Integration Theory.

PMA-R: Primary Mental Abilities test: Reasoning scale.

PMA-V: Primary Mental Abilities test: Vocabulary scale.

PS: Processing Speed.

Q: Question.

QBI: qball imaging.

RAPM: Raven Advanced Progressive Matrices Test.

RD: Radial diffusivity.

RF: Radiofrequency.

RMSEA: Root Mean Square Error of Approximation.

ROI: Region of interest.

SAT: The Scholastic Achievement Test.

SBM: Surface-Based Morphometry.

SC: Structural Connectivity.

SPM: Statistical Parametric Mapping.

STM: Short-Term Memory.

TE: Echo Time.

TR: Repetition Time.

T1: Relax Longitudinally Time.

T2: Transverse Relaxation Time.

UPD: Executive Updating.

VBM: Voxel-Based Morphometry.

VPR: Verbal-Perceptual-Rotational model.

WM: White Matter.

WMC: Working Memory Capacity.

List of Figures

Chapter 1: Intelligence and Cognitive Processes

Chapter 2: Attempts to raise intelligence

Chapter 3: Neuroimaging Framework for Studying Brain Plasticity

Chapter 5: Methodology and Basic Design

Chapter 6: Training performance (Q1)

Chapter 7: Gray matter responsiveness and n-back training (Q2.1)

Chapter 8: Structural connectivity and n-back training (Q2.2)

Chapter 9: Follow-up (Q3 & Q4)

Chapter 10: General Discussion

List of Tables

Prefacio

"Todo hombre puede ser, si se lo propone, escultor de su propio cerebro"

Santiago Ramón y Cajal (1897)
Reglas y consejos sobre investigación científica.

El estudio de los efectos del entrenamiento cognitivo posee una rabiosa actualidad en ciencia. Tan diversos son los tipos de entrenamiento como los efectos. Aunque no es un campo novedoso, sigue habiendo muchas incógnitas sobre los efectos potenciales de *ejercitar nuestra mente*. La palabra 'ejercitar' no es casual, ya que a menudo se supone que el entrenamiento cognitivo tiene un paralelismo directo con el entrenamiento físico. Pero, *¿podemos mejorar nuestras capacidades (abilities) mediante el entrenamiento cognitivo*?

Esta pregunta es un clásico en Psicología Diferencial. A finales de los 60 se desarrollaron numerosos programas para mejorar las capacidades de los niños en situación de desventaja. Proyectos como Head Start, Milwakee o Abecedarian son ejemplos paradigmáticos. Sin embargo, la ausencia de consenso sobre su efecto hizo que los científicos abandonasen prácticamente el campo.

Pero la publicación del artículo del equipo de Susanne M. Jaeggi en 2008 en la revista PNAS supuso una revolución (Jaeggi y cols., 2008). En este trabajo se concluía que la mejora de la memoria de trabajo, tras completar un exigente entrenamiento basado en la tarea n-back, impactaba en un aumento de la inteligencia fluida.

A partir de este hallazgo comienza una nueva edad de oro para este tipo de estudios. Numerosos laboratorios se pusieron manos a la obra para replicar ese impactante resultado, pero los productos han sido heterogéneos. De los tres meta-análisis publicados dos concluyen que el entrenamiento de la memoria de trabajo no tiene efectos estables sobre la inteligencia (Melby-Lervåg y Hulme, 2013; Schwaighofer y cols., 2015). Sin embargo, el meta-análisis de Au y cols. (2014), centrado en la n-back, detectó un efecto débil, pero significativo, sobre la inteligencia (3.6 puntos de CI). Según el análisis de Herrnstein y Murray (1994) un efecto de esa magnitud poseería un enorme efecto en la población.

En paralelo a los estudios psicológicos sobre el entrenamiento cognitivo, se desarrollaron investigaciones sobre el dinamismo del cerebro humano. La idea de un cerebro estático, en el que, por ejemplo, no se producen neuronas después de la infancia, ha sido abandonada (Pascual-Leone y cols., 2005). No sólo se han estudiado los cambios madurativos, sino que también se ha explorado el efecto del entrenamiento (motor y cognitivo) sobre la estructura cerebral.

¿Estará en lo correcto Ramón y Cajal?
¿Podemos cincelar nuestro cerebro?

La presente tesis doctoral explora la respuesta a esa pregunta en las siguientes fases:

Fase 1. Primera evaluación psicológica (pretest).
Fase 2. Selección de candidatos: grupo control y de entrenamiento.
Fase 3. Adquisición de imágenes MRI.
Fase 4. Entrenamiento cognitivo (3 meses).
Fase 5. Adquisición de imágenes MRI y segunda evaluación psicológica (posttest).
Fase 6. Seguimiento.

En este trabajo se estudian los efectos del entrenamiento adaptativo en la n-back dual sobre una serie de constructos (estimados a partir de tres medidas cada uno): inteligencia fluida, inteligencia cristalizada, memoria de trabajo y control atencional.

A nivel biológico se estudia el efecto sobre la estructura cerebral, aplicando técnicas de morfometría basada en superficie (SBM) porque permite estudiar cambios en el grosor cortical (cortical thickness, CT) y el área de superficie cortical (cortical surface area, CSA). Además, se estudia el efecto del programa sobre las conexiones (materia blanca) que unen regiones (predefinidas) de la corteza.

Para evitar los problemas metodológicos denunciados por Thomas y Baker (2013a) aquí se estudia la interacción. Por tanto, los cambios psicológicos y biológicos del pretest al posttest (TIEMPO) del grupo de entrenamiento se han comparado con los cambios ocurridos en un grupo control (GRUPO).

Por último, se ha comprobado (a) si las mejoras conductuales observadas durante el entrenamiento se mantienen en el tiempo y (b) la posible transferencia al rendimiento académico.

Esta tesis doctoral se ha desarrollado en colaboración con laboratorios de Canadá (MNI) y Estados Unidos (LONI), en los que el doctorando disfrutó de una estancia gracias a la financiación del gobierno de España (beca FPI: BES-2011-043527). Esa colaboración ha contribuido a mejorar el producto presentado a continuación.

Finalmente se presentan los bloques en los que se divide este informe.

El primer bloque es conceptual: (a) definición de conceptos, (b) revisión de la investigación sobre la mejora de la inteligencia después del entrenamiento cognitivo, (c) cambios cerebrales (madurativos y evocados), y (d) hipótesis sobre las preguntas (P) planteadas en la investigación:

P1: ¿Cómo fue el rendimiento durante el programa de entrenamiento?
P2: ¿Se producen cambios en las estructuras del cerebro debido al entrenamiento?
P3: ¿Se mantiene el nivel de rendimiento alcanzado un año después?
P4: ¿Hay diferencias entre el grupo control y el grupo de entrenamiento en el rendimiento académico tras la finalización del programa?

El segundo bloque presenta el estudio experimental y los resultados observados. En el primer capítulo se detalla el diseño, mientras que los siguientes capítulos se centran en los efectos a nivel psicológico y biológico. Los resultados del estudio de seguimiento se presentan en el capítulo nueve. Las conclusiones generales cierran el informe. [1]

[1] Parte de esta tesis doctoral ha sido publicada (ver Colom et al., 2013a)

PART I: Theory

We live in a society exquisitely dependent on science and technology,

in which hardly anyone knows anything about science and technology.

- Carl Sagan (1990)
Why we need to understand science.

In Part I. the conceptual framework, aims and the main research questions are defined. The general purpose of this work was to study the possibility of raising intelligence through an adaptive cognitive program based on the n-back task. In addition to psychological analyses, neuroimaging methods were employed for studying brain indices related with cognitive performance. Finally, training effects were checked in a follow-up study.

Chapter one briefly reviews the concept of intelligence and related cognitive processes. The second chapter describes attempts to raise cognitive ability. The third chapter introduces basic concepts and methods in neuroimaging for obtaining the biological indices considered here. Finally, the framework, key aims and research hypotheses are specified in the last chapter.

CHAPTER 1: Intelligence and Cognitive Processes.

1.1. Introduction.

Intelligence and cognitive processes are main topics of scientific psychology. The study of intelligence has predominantly been carried out using correlational approaches, whereas experimental approaches have typically been used to study cognitive processes. The experimental method is based on the manipulation of independent variables to detect their effects on a predefined dependent variable. In contrast, correlational methodology analyzes patterns of co-variation among diverse psychological measures. Both disciplines of scientific psychology have traditionally run in parallel. Fortunately, in the last years, both approaches are being combined in the scientific study of human behavior.

Human intelligence is maybe the psychological factor most frequently analyzed at the intersection of both approaches (Wilhelm & Engle, 2005). Hunt (2011) argued that some cognitive processes might be considered as mechanisms of intelligence and that therefore, the combination of both topics is highly relevant. In this chapter, the definition of intelligence, its measurement and its structure are described. Afterwards, the mechanisms of intelligence (cognitive processes) are discussed. The understanding of cognitive processes and human intelligence is crucial for understanding this work because it focuses on the effects of an adaptive cognitive training program on both.

1. 2. Definition.

"*The study of intelligence is not an isolated academic topic; our intelligence has social consequences*" (Hunt, 2011, p.1).

The relevance of intelligence for fields such as education, economic development or health, have generated a broad interest beyond Psychology, due to intelligence predicting a large set of social outcomes (Neisser et al., 1996, Nisbett et al., 2012). Intelligence is a psychological construct, but also a brain property that contributes to everyday life behavior (Johnson, 2013). At the turn of the 20th century, the scientific community provided solid consensus regarding the definition of intelligence:

"Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings-"catching on," "making sense" of things, or "figuring out" what to do" (Gottfredson et al., 1997, p. 13).

This definition is highly consistent with the definition of intelligence provided by the American Psychological Association (APA).

"*Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought*" (Neisser et al., 1996, p. 77).

People differ in their ability to reason, to solve every day life problems, and to learn from formal and informal experience. These differences can be reliably measured.

1. 3. Measuring Intelligence.

The first formal approach for measuring intelligence is usually attributed to Alfred Binet, who devised tests to distinguish mentally retarded children from those with behavior problems (Neisser et al., 1996; Jensen, 1998). Binet used the mental age concept to classify children who were entering school.

Intelligence tests measure performance in a variety of tasks (Gottfredson, 1997; Hunt, 2011; Jensen, 1998; Johnson, 2013; Nisbett et al., 2012). Intelligence tests are among the most accurate of all psychological tests and assessments (Gottfredson, 1997), and intelligence test scores are highly stable. In this regard, Deary, Whalley, Lemmon, Crawford and Starr (2000) reported a test-retest correlation of .73 for a time gap of 66 years.

Intelligence tests scores are usually transformed to the Intelligence Quotient (IQ) scale. This scale (with a mean of 100 and a standard deviation of 15) has some important features. IQ scores are normally distributed and this has two relevant implications:

(a) Approximately two-thirds of the scores are in the range 85 – 115.

(b) Five percent of the scores are higher than 125 (or lower to 75). Scores above 135 or inferior to 65 are only found in one percent of the population.

IQ scores are standardized according an age group. Thus, IQ measures a person's relative performance on an intelligence test, compared to the performance of people in an appropriately chosen comparison group of the same age. Note that a score of 130 does not mean that the individual is 30% smarter than someone with an IQ score of 100. A score of 130 situates the person in the highest 2% of the proper reference group. Consequently, the IQ scale is not a ratio scale such as other measures like weight, liter, or kilogram (Haier, 2014).

Nevertheless, intelligence tests are highly relevant because they are used for many purposes, such as selection, diagnosis, and evaluation (Neisser et al., 1996). However, there are different types of intelligence tests. There are tests that only use one type of item (e.g. Raven's Progressive Matrix test, Raven, 1988), while others use a combination of items (e.g. The Wechsler intelligence scales, Wechsler, 2003, 2008). There is high variability in the types of items: verbal, logical, numerical, spatial, etc. Moreover, there are intelligence tests requiring specific cultural knowledge (e.g. vocabulary) and tests based on simple, universal concepts (many/few, open/closed, up/down). Nevertheless, all intelligence tests measure the same general construct (Gottfredson, 1997), which is demonstrated by the *positive manifold* phenomenon; performance on varied intelligence tests is positively correlated, which leads to the extraction of the general factor of intelligence (*g*) (Carroll, 1993; Jensen, 1998; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Spearman, 1904).

1. 4. Structure of Intelligence: psychometric models.

Structural models analyze the correlation matrices defined by varied cognitive tasks. These models assume that variations in cognitive performance can be summarized by a number of basic cognitive dimensions, such as general reasoning, verbal facility, and visual-spatial reasoning (Colom, 1998; Jensen, 1998; Juan-Espinosa, 1997; Andrés-Pueyo, 1996). As noted above, scores obtained on different intelligence tests are positively correlated, which allows data reduction and the extraction of latent factors at different levels of generality. Different models have been proposed to summarize the structure of intelligence. Examples of some models are presented here:

(1) The theory of General Intelligence (*g*). This model was proposed by Spearman (1904). He observed that the performance of school subjects in a set of tests was positively correlated. Based on this evidence, he proposed that individual differences in cognitive performance were largely due to individual differences on a single dimension of mental competence, the general factor of intelligence (*g*). Also, a group of specific factors were included in the model (Hunt, 2011) (Figure 1.1).

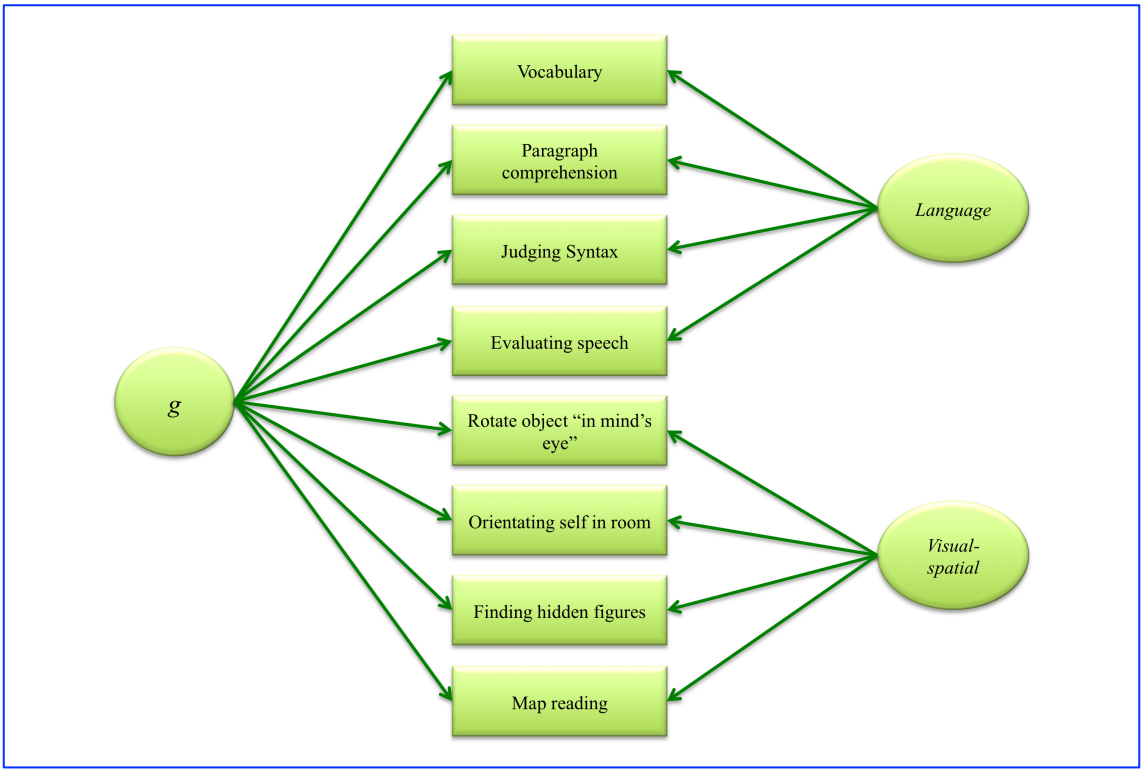


Figure 1.1. Spearman's model. The two group factors are uncorrelated (adapted from Hunt, 2011). Circles are latent variables while rectangles represent manifest variables.

(2) The Three-Stratum Model (CHC model). This model was originally developed by Raymond Cattell (Cattell, 1971, 1987) and John Horn (Horn, 1985), and modified by John Carroll (Carroll, 1993). Cattell and Horn considered that Spearman's model gave inadequate weight to group factors. They defined a model with *two stratums* without a general factor of intelligence. The first-stratum was defined by narrow abilities measured by different tests (vocabulary, calculus, logical reasoning, etc.). The second-stratum comprised broad abilities. The number of factors at this level was diverse, although two distinguishable factors, albeit highly correlated, were especially salient: fluid intelligence (Gf) and crystallized intelligence (Gc). Gf was defined as the ability to

deal with new and unusual problems, which highly depended on brain maturation. In contrast, Gc was the ability to apply previously acquired knowledge, so it mainly depended on experience (Cattell, 1987). Li et al. (2004), among others, showed that the development of Gf and Gc is highly different. Fluid abilities correlated more with sensory processing while crystallized abilities are more related with socio-biographical predictors. This outcome supports the idea that these two facets of intelligence differentially reflect neurobiological and sociocultural influences on intellectual development (see Figure 1.2).

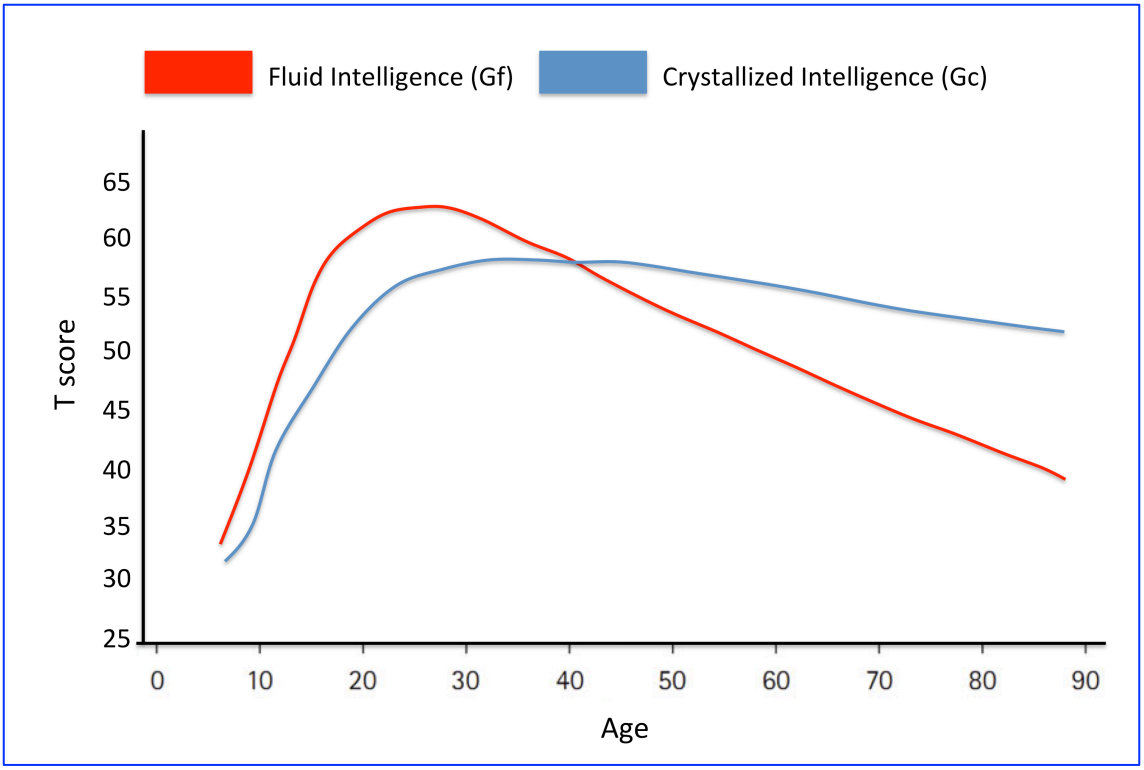


Figure 1.2. Developmental changes of fluid intelligence (Gf) and crystallized intelligence (Gc) scores across life span. Adapted from Li et al. (2004). T scores (mean = 50, standard deviation = 10).

The seminal study by Carroll was based on the reanalysis of over 450 data sets. The main findings supported the Gf-Gc model. However, second-stratum abilities were correlated, which led to the general factor of intelligence (McGrew, 2009). Further, more factors in the second stratum were proposed (see Figure 1.3). Gf and Gc are still at the core of the model. Indeed, Gf is identified with *g* in many studies (see Colom, 1998). The ability to deal with visual images (visual-spatial ability; Gv) is also important in many contexts (Gustafsson, 1984), therefore Gv is also frequently study.

Note that factors in Figure 1.3 are depicted in order of relevance to explain the structure of intelligence.

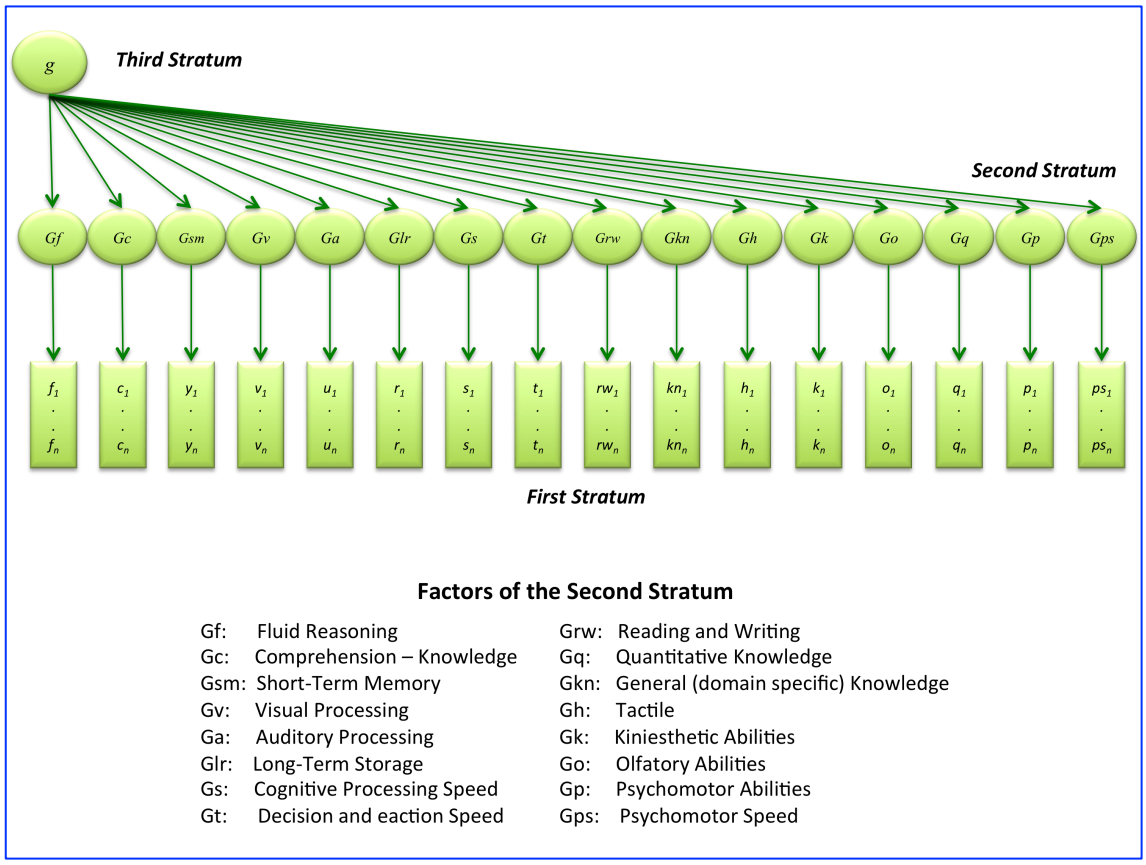


Figure 1.3. CHC model; adapted from McGrew (2009). Circles are latent variables while rectangles represent manifest variables.

(3) The verbal – perceptual – rotational model (*g*-VPR model). Johnson and Bouchard (2005) proposed a model with *four-strata*. The *first-stratum* included primary traits/factors evaluated by individual tests. The *second-stratum* consisted of broader but still narrow abilities (word fluency, verbal comprehension). The *third-stratum* contained three factors: verbal (V), perceptual static (P) and perceptual rotation (R). Finally, the *four-stratum* comprised a single factor, identified with the general factor of intelligence (*g*).

The VPR model identified Gc with the verbal factor, Gf with *g*, and Gv was divided into perceptual ability (focused in static images) and the ability to conduct mental manipulations of visual objects (rotation). It is argued that this model is more plausible at the biological level than the CHC model. It has also shown a better psychometric fit (Figure 1.4).

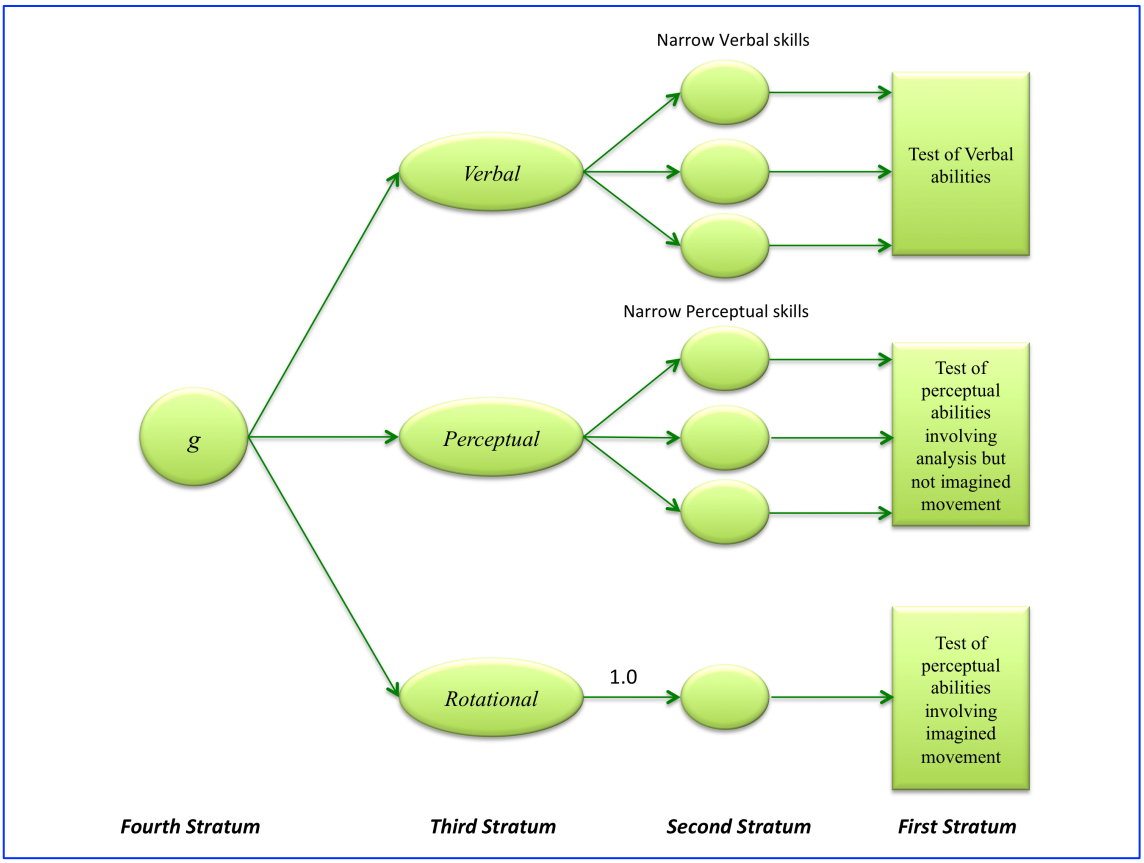


Figure 1.4. VPR model; adapted from Johnson and Bouchard (2005). Circles are latent variables while rectangles are observed variables.

(4) Conclusion. Both the CHC and the VPR models are proper summaries of the structure of the intelligence construct (Hunt, 2011). Both agree in the hierarchical structure of human intelligence, where a higher order factor representing general intelligence (*g*) is located at the top of the hierarchy, indicating the level of difficulty that an individual can handle in cognitively demanding situations.

There are some theories suggesting that intelligence must be viewed as a collection of separate cognitive abilities, despite the empirical evidence supporting the relevance of the *g* factor (Cattell, 1987; Gardner, 2004; Guilford, 1988; Sternberg, 1988; Thurstone, 1938). Nevertheless, as discussed by Colom and Thompson (2011), among others, the conflict is not real. The *g* factor emerges when (a) reliable and valid measures of intelligence are used, and (b) restriction of range and measurement errors is corrected. The positive manifold is inevitable when representative samples of the general population are considered.

1. 5. Mechanics of Intelligence: cognitive processes.

Psychometric theories describe the structure of intelligence by summarizing the variations in tests' performance across populations (Hunt, 2011). Likewise, they are useful in predicting the performance of people on diverse problems. But they cannot explain the cognitive process or processes behind the score in a test or factor (Deary, Austin, & Caryl, 2000). Answering the question 'How are the problems solved?' requires a different approach based on information-processing models.

The study of cognition has been historically separated from the study of individual differences in intelligence. But this changed in the early 70s. Cognitive processes are hypothetical constructs describing how people apprehend, discriminate, select and attend to certain aspects of stimuli. These features impinge on the sensorium to form internal representations that can be mentally manipulated, transformed, related to previous internal representations, stored in memory (short-term or long term memory), and later retrieved from storage to govern the person's decisions and behavior in a particular situation (Jensen, 1998).

There are different theories or models of information processing, but two aspects are included in all models: (a) the idea that cognitive processing occurs in stages, either serially or in parallel; and (b) that the occurrence is in real time, that is, each step in the process requires a certain time interval. All these processes are supported by structural and physiological properties of the brain. A model of how this exchange of information takes places is called a *cognitive architecture*. Figure 1.5 depicts a representation of the basic cognitive processes included in a general model of human cognition (Hunt, 2011).

These are the components of the model:

Sensory system. The information from the environment is sensed through different sensory systems (visual, auditory, tactile, etc.), and then must be discriminated and encoded.

Cognitive system. Then, the information is classified into higher-order categories by activating related information stored in *long – term memory* (LTM). Items in LTM exist in various states of activation, depending on how frequently and how recently they have been attended to. Next, the information retrieved from LTM, together with the

information coded from the environment, is located in *working memory* (Baddeley, 1986), where the cognitive processing required for the current task is carried out. Working memory comprises the functions of focusing attention, conscious rehearsal and transformation and mental manipulation of information (Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004). Working memory is considered central for the human information processing system (Kyllonen & Christal, 1990).

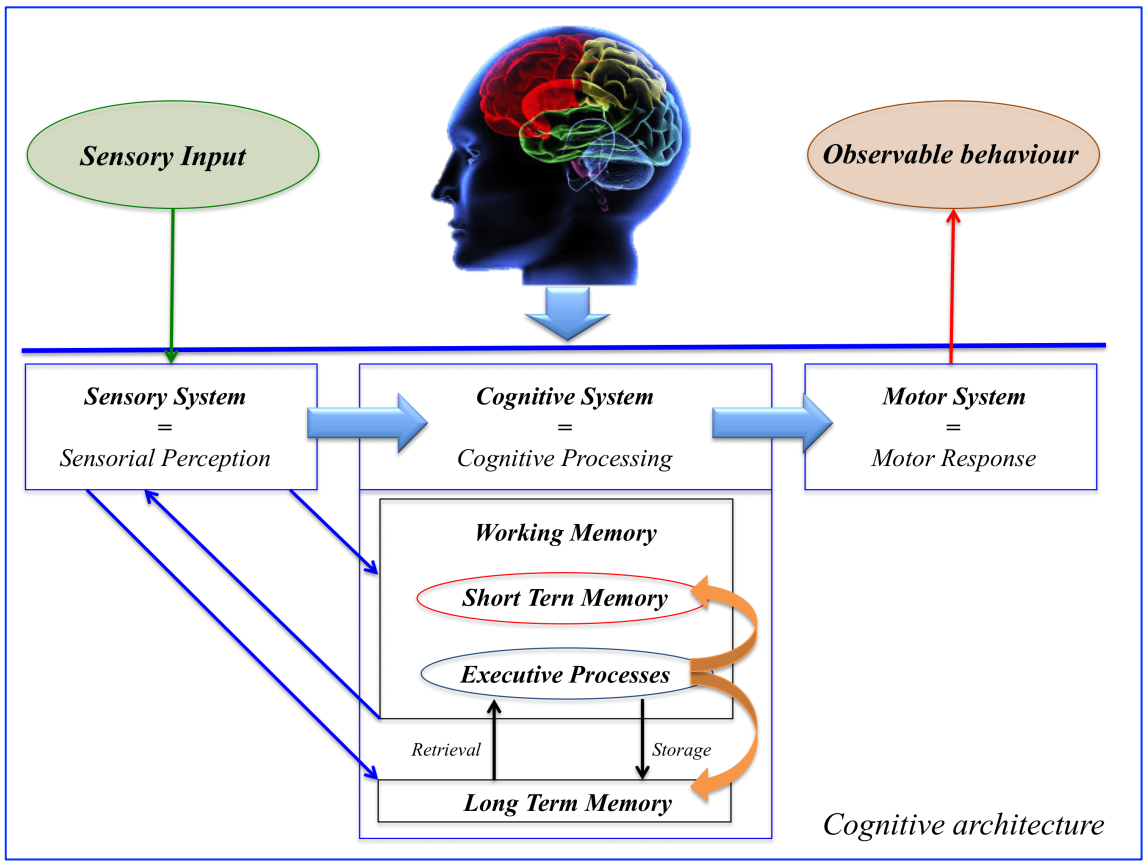


Figure 1.5. Basic components included in a general model of human cognition (adapted from Martínez, 2014).

The working memory system requires temporary storage (*short-term memory*, STM) and further information processing (Baddeley, 2007). The internal representation is updated under the influence of the new incoming information and relevant information that have been stored in LTM. The *executive processes* (EP) are in charge of this. They can be divided in (1) *attention control* (ATT), defined as the ability to maintain relevant mental representations in a highly active state while suppressing the distracting one (Engle, Kane, & Tuholski, 1999); and (2) executive functioning that can be decomposed in more specific processes such as *inhibition, shifting, and updating* (Miyake, Friedman,

Emerson, Witzki, & Howerter, 2000), which implicate the suppression of automatic responses, switching between tasks, and the on-line addition or subtraction of information from the working memory system respectively (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008). The 'storage' section of working memory may be divided into different modality-specific sections where the linguistic, auditory-nonlinguistic and spatial/visual information are temporarily represented and kept active. Information in working memory is consolidated in LTM. Also, *Processing speed* (PS) could limit working memory capacity, because processing operations (e.g. ATT, updating…) take time and the storage component (equivalent to STM) has a limited capacity. In this regard, Burgaleta and Colom (2008) report results suggesting that STM and PS are the only components predicting intelligence performance.

Motor system. This system is responsible for the motor response (speaking or making a psychological movement) demanded by the ongoing cognitive processing in working memory.

There are remarkable individual differences in all these steps, but mainly in the processes occurring in the cognitive system. There are numerous studies exploring the relationships between intelligence differences, as assessed by standardized tests, and performance differences on a range of cognitive tasks. To date, the main conclusion is that working memory is the best predictor of variability in intelligence (e.g., Colom, et al., 2004, 2008; Colom, Abad, Rebollo, & Shih, 2005; Cowan, 2005; Kane et al., 2004; Kyllonen & Christall, 1990; Unsworth, Redick, Heitz, Broadway, & Engle, 2009; for a review, see Ackerman, Beier, & Boyle, 2005). Nevertheless, there is a lack of consensus regarding which working memory components are mainly responsible for this large relationship (Deary, 2000). Some theories underline processing speed (Jensen, 2006), while other models appeal to attention scope (Cowan, 2005; Engle, Tuholski, Laughlin, & Conway, 1999), dual mechanisms for information maintenance and retrieval (Oberauer, Süß, Wilhelm, & Wittman, 2003; Unsworth, et al., 2009), or simple short-term storage (Colom, et al., 2004, 2005, 2008; Martínez et al., 2011).

In an exhaustive study reported by Martínez et al., (2011) limitations of previous studies were addressed: (a) limited range of tasks to measure the constructs of interest, (b) arguable scores derived from these tasks, (c) analyses of raw correlations instead of latent factors, and (d) use of small and improperly recruited samples (Unsworth et al.,

2009). Eight psychological constructs were considered: fluid intelligence (Gf), crystallized intelligence (Gc), and visual-spatial intelligence (Gv), short-term memory (STM), working memory capacity (WMC), executive updating (UPD), attention control (ATT), and processing speed (PS). Results indicated that individual differences in Gf could be largely accounted for by basic (general, but not unitary) mental processes underlying memory span, namely, encoding, maintenance, and retrieval, since short-term storage is the cognitive component common to the three span measures (STM, WMC and UPD; see Colom, et al., 2006, 2008; Hornung, Brunner, Reuter, & Martin, 2011; Krumm et al., 2009) (Figure 1.6).

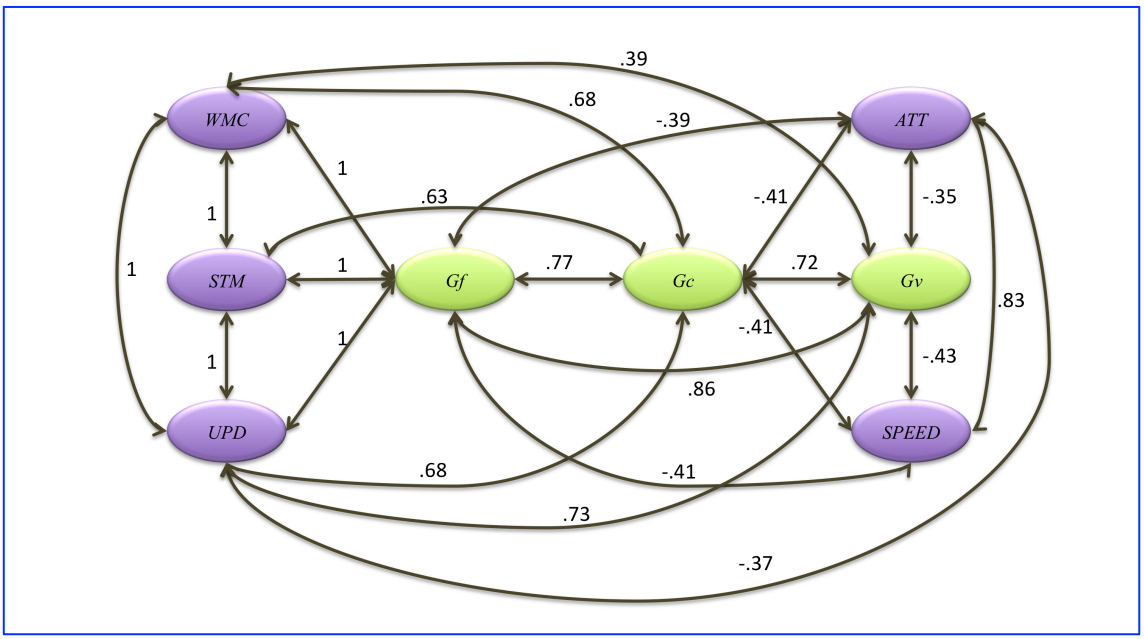


Figure 1.6. Confirmatory factor analysis comprising all the relevant constructs in the study by Martínez et al., (2011). WMC = working memory capacity, STM = short-term memory, UPD = executive updating, ATT = attention control, Gf = fluid intelligence, Gc = crystallized intelligence, Gv = visual-spatial intelligence. Violet circles depict constructs from experimental cognitive, while green circles depict constructs from correlate approach. Only significant correlations are represented. The simultaneous relationships among STM, UPD, WM, and Gf are fixed to be equal and with a value of 1 (adapted from Martínez et al., 2011).

These results are highly consistent with new findings, showing that fluid intelligence might be improved by training aimed at increasing memory span (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). As discussed by Colom et al. (2010a), memory span factors and intelligence share relevant cognitive mechanisms. Also, neuroimaging studies have shown relevant common structural and functional networks underlying intelligence and

memory span (Colom, Jung, & Haier, 2007; Gray, Chabris, & Braver, 2003). These results will be discussed in more detail in Chapter 3.

Note, however, that these findings do not mean that other processes are irrelevant for predicting intelligence when there are considered in isolation. The important point is that they fail to add significant information in answering the question regarding the basic cognitive processes underlying fluid reasoning when several diverse factors/functions are simultaneously considered and compete in the predictive nomological network.

1. 7. Conclusions.

Human intelligence is a general mental ability for reasoning, problem solving, and efficient learning. It can be identified with both a higher-order mental ability and several related abilities and skills that can be properly measured by standardized tests. Individual differences in intelligence (a) become more salient as the cognitive complexity of the situation becomes greater, (b) are stable over time, (c) and revealed in many disparate social situations. Understanding the mechanics of intelligence requires studying cognitive processes. In this regard, the role of working memory capacity must be underscored. Indeed, it has been shown that the general factor of intelligence (*g*) and working memory capacity are very largely correlated. The fact that (a) intelligence, as assessed by standardized tests, predicts a large set of social outcomes, and that (b) intelligence is strongly correlated with working memory capacity, opens the door to the possibility of increasing intelligence via improving the capacity of the working memory system. This will be the main topic of the next chapter.

CHAPTER 2: Attempts to raise intelligence.

2.1. Introduction.

In this study, a cognitive program based in n-back task was used in an attempt to raise intelligence. A brief description about the main concepts of intelligence and some related cognitive processes were described in the previous chapter. These ideas are highly important to understand the findings from new cognitive programs trying to improve intelligence. This chapter begins by discussing the benefits of an improvement in intelligence. Then, it includes a brief description of the first studies dedicated to raise intelligence, as well as, the new programs. These programs can be classified in three different approaches: pharmacological/nutritional, educational, and psychological. Especially relevant here is the psychological approach, which studies the improvement in cognitive processes and their transfer to intelligence measures.

2. 2. Benefits of improving intelligence.

Intelligence test scores predict many real world phenomena and have well-validated practical uses (Deary, 2012; Deary, Penke, & Johnson, 2010; Gottfredson, 1997; Nisbett et al., 2012). For example, intelligence is the best single predictor of academic performance (Kuncel, Hezlett, & Ones, 2001). Indeed, there is an important relationship between intelligence and educational level/attainment increasing 2.1 IQ points per year of education (Abad, Sorrell, Román & Colom, in press; Ceci, 1991; Gustafsson, 2001). In other contexts, such as job performance, social status and income, intelligence is also the best predictor (Herrnstein & Murray, 1994; Schmidt & Hunter, 2004), even higher than other factors, like emotional intelligence (Amelang & Steinmayr, 2006; Brody, 2004; Matthews, Roberts, & Zeidner, 2004). Emotional intelligence is highly employed in job contexts, however its measure has less validity and reliability than traditional intelligence scores (Amelang & Steinmayr, 2006; Brody, 2004; Matthews et al., 2004). Also, general intelligence is related to social class, health knowledge, and health literacy (Gottfredson, 2004). Moreover, intelligence is a predictor of longevity (Whalley & Deary, 2001). These are only some examples of the relevance of intelligence in different social contexts.

Herrnstein and Murray (1994) reported a simulation, using the *National Longitudinal Study on Youth* (NLSY), to test the benefits of raising the mean level of general cognitive ability by just three IQ points (from 100 to 103). The NLSY is a large, nationally representative sample of American youths who were assessed from 14 to 22 years (originally 12,686 persons). The study began in 1979, and has been followed ever since. These were the results of the simulation:

"The poverty rate falls by 25 percent. So does the proportion of males ever interviewed in jail. High school dropouts fall by 28 percent. Children living without their parents fall by 20 percent. Welfare recipiency, both temporary and chronic, falls by 18 percent. […]. The incidence of low-weight births drops by 12 percent. Children in the bottom decile of home environments drop by 13 percent. Children who live in poverty for the first three years of their lives drop by 20 percent" (Herrnstein & Murray, 1994, p.367).

3 IQ points are not particularly relevant at an individual level, but the effect at a population level is socially relevant.

The high heritability of intelligence is widely accepted, with values ranging from .40 to .80 (Plomin, de Fries, McClearn, & McGuffin, 2008; Posthuma et al., 2002; Pol et al., 2006; Chiang et al., 2009, 2012). Nevertheless, this does not imply that intelligence is insensitive to the environment. Genes speak about potential, but the environment modulates the realization of this potential (Au et al., 2014; Hunt, 2011). Improving the physical, social, and mental environments would promote intelligence development (Hunt, 2012a).

2. 3. Goals of studies focusing on improving intelligence.

"*As a construct, the g factor* (general intelligence) *can be represented with varying degrees of convenience, efficiency, and validity by a wide variety of vehicles (psychometric tests, laboratory techniques, physiological indices) that yield measurements that have different scale properties*" (Jensen, 1998, p. 306).

The difference between constructs, vehicles and measurements is crucial for a proper understanding of what it means, at the construct level, to raise intelligence scores. IQ scores estimate intelligence levels, but changes in IQ scores (measurement) may or may not involve changes at the construct level (intelligence) because distinguishable sources

of variance are implicated (Jensen, 1998). Test scores change by simple test-retest effects (Anastasi, 1934; Estrada, Ferrer, Abad, Román, & Colom, 2015; Jensen, 1998; Reeve & Lam, 2005) yet it has been observed that these practice effects are hollow (i.e. they don't involve an increase in general intelligence) (Jensen, 1998; te Nijenhuis et al., 2007). Specific skills tapped by the measures can show substantial increments, but the general construct tapped by these measures (*g*) remains unaltered.

It's important to note that the ability to solve problems, which is defined in many contexts as Gf, has an important biological substrate. Therefore, the raw scores of Gf tests are highly malleable across the life span due to maturational development. For example, a ten year old child could solve more complex problems than he could when he was five years old. However, independently of these improvements due to maturation, the IQ score is highly stable over time (Deary et al., 2000) i.e. the position with respect to the age-matched group is highly unchanging. Figure 2.1 shows an example of a hypothetical participant assessed in Gf across her/his life. If this imaginary participant completes a cognitive training program, and the results are satisfactory, his/her intelligence score would improve. In other words, his/her percentile rank would improve with respect to his/her normative group.

We can do a direct analogy between physical training and cognitive training. The benefits of physical training are visible and the improvements are maintained for some time. But the effects would likely not be visible if the training sessions were too short or people didn't keep training for a sufficiently extended period of time. Moreover, effects may disappear if training stops. In the same way, intelligence improvement programs do not work if they are applied only for a short time (Vallejo-Nájera & Colom, 2004). This being said, the main goal of cognitive training programs is to raise the raw test scores with respect to a normative group or, stated differently, to raise intelligence. However, these increments must be higher than a simple test-retest effect. Therefore, to speak about increment in intelligence scores, the results of cognitive training must be compared with a control group to control for this effect. However, the benefits of cognitive training programs may disappear once the program stops.

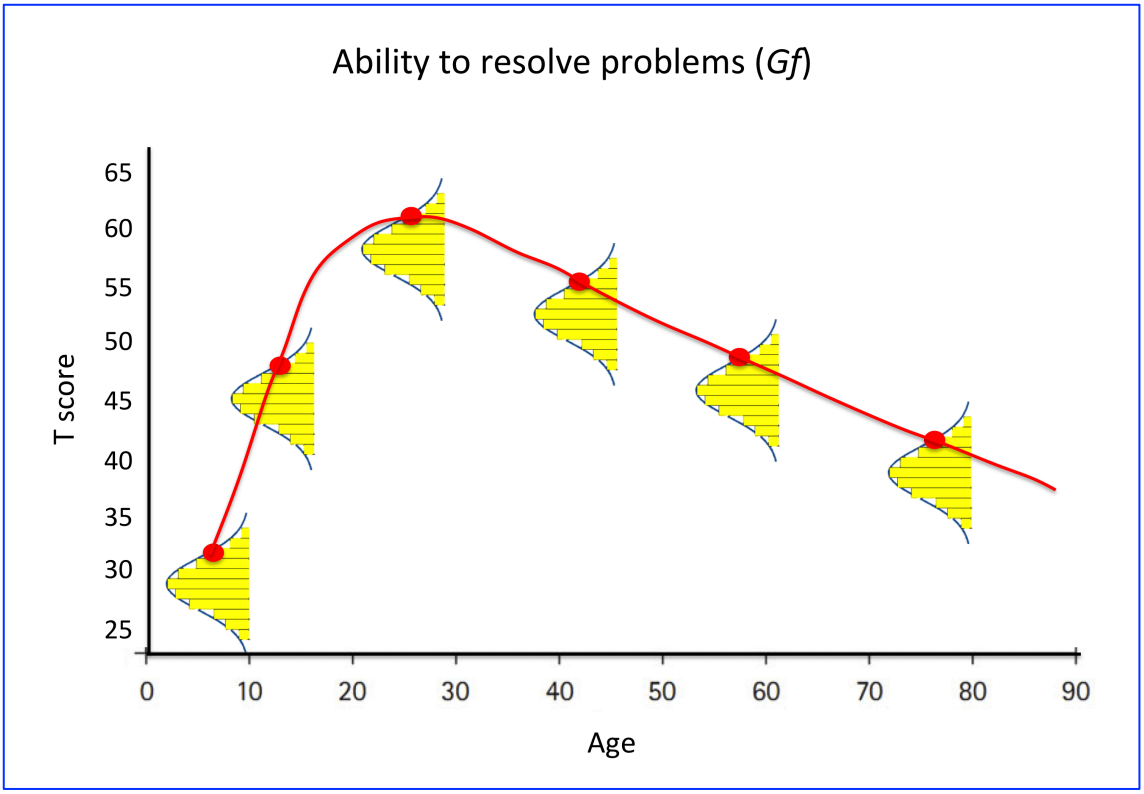


Figure 2.1. Evolution of the ability to resolve problems across life span. T score is a scale with mean 50 and standard deviation 10. Red dots show the IQ scores of a participant inside their normative group. Gf = fluid intelligence. Gf trajectory is adapted from Li et al. (2004).

2. 4. Pioneer programs.

Three programs will be described in this section: (a) Head Start, (b) Milwaukee and (c) Abecedarian. These studies were carried out with children in impoverished environments and with a high risk for mental retardation. In all three cases, the interventions started very early in life.

(1) Head Start was launched in 1964 by the USA government (see Zigler & Muenchow, 1992). The program was designed to improve learning, social skills and health status in poor preschoolers and promoting regular schooling (Jensen, 1998; Herrnstein & Murray, 1994). The intervention involved short periods of time (from a few months to two years). After hundreds of studies, the general conclusion was that the observed cognitive benefits eventually faded after the end of the program (McKey, 1985; Zigler & Muenchow, 1992; see Figure 2.2).

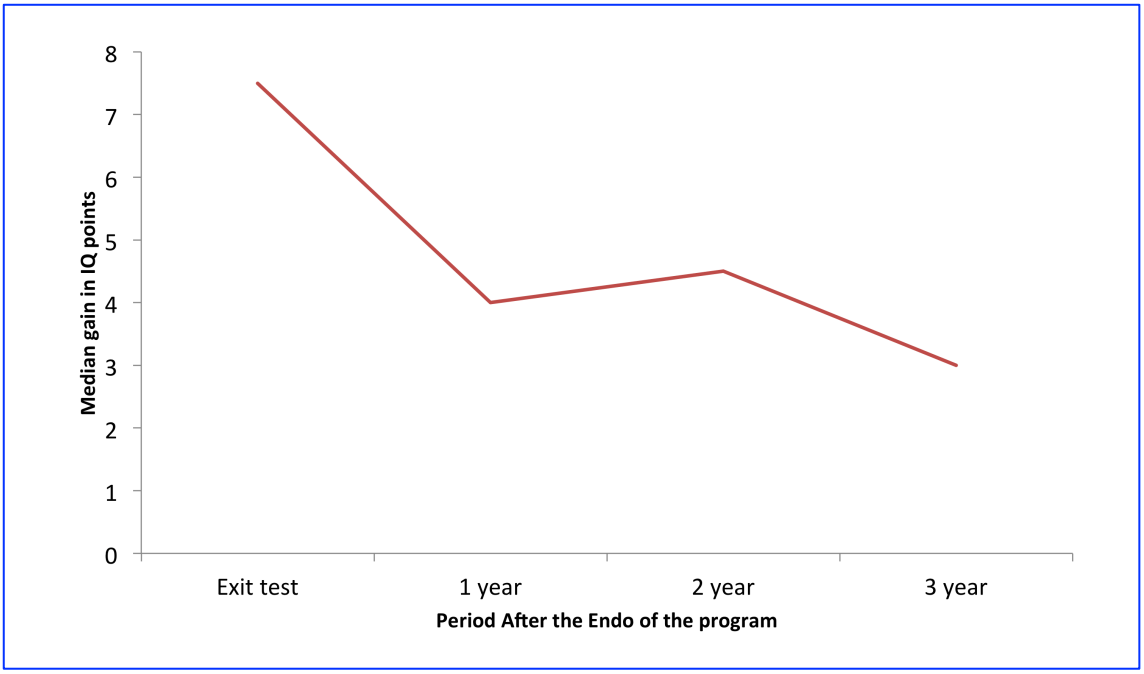


Figure 2.2. IQ gains attribute to Head Start Program (adapted from Herrnstein & Murray, 1994)

(2) The Milwaukee Project was launched in 1966 (see Garber, 1988). The program was implemented to help children at high risk for mental retardation because they were born in poor environments from mothers whose IQs were seventy-five or below (Jensen, 1998). Children in the training group (N = 17) went 5 times per week to an 'Infant Stimulation Center', where they completed several activities. At the end of the program, the difference between the training and control groups (N = 18) was equivalent to 32 IQ points. But, after eight years from the end of the program, the difference decreased to 10 IQ points. This reduction was especially higher on measures departing from the training exercises (Jensen, 1989; Locurto, 1991). Furthermore, there were no differences in scholastic achievement. In conclusion, only near-transfer effects were found.

(3) The Abecedarian project (see Ramey, MacPhee, & Yeates, 1982) resembles the Milwaukee project: the targets were infants at high risk for mental retardation. The intervention comprised training sessions 5 days per week, fifteen weeks per year during 5 years. One important difference between Milwaukee and this project is that more than one criterion was critical for selection. Therefore, children with high IQ mothers participated in this project in contrast to the Milwaukee project. The training and control groups were assessed with appropriate intelligence test (e.g. Stanford-Binet and

Wechsler scales) (Jensen, 1998). The overall average difference between groups was of 7.8 IQ points (training > control). In follow-up studies (after 10 years) the difference was 5 IQ points and scholastic achievement was also greater in the training group (Jensen, 1998). Only 13 percent of the experimental children had IQs of less than 85, compared to 44 percent of the controls (Ramey, 1992). However, the debate is still active because the training and control groups were not comparable in their intellectual prospects at birth (Herrnstein & Murray, 1994).

The main outcome of these programs was a big difference between the training and control groups at the end of the program. However, this difference progressively decreased which suggests that (a) only near-transfer effects were observed, (b) the environments of the children in the training group did not remain enriched (Nisbett et al., 2012) and (c) these programs were designed for children with a priori low intelligence (see Table 2.1).

Table 2.1. Results of pioneer training programs. Differences between training and control group (training > control) are depicted here.

Study	*Duration*	*Gain at end*	*Follow-up studies*	*Reduction IQ per year*
Head Start project	2 months-2 years	7.5 IQ points	3 IQ points (After 3 years)	1.5 IQ points
The Milwaukee project	6 years	32 IQ points	10 IQ points (After 8 years)	2.75 IQ points
The Abecedarian project	5 years	7.8 IQ points	5 IQ points (After 10 years)	0.28 IQ points

The highest and lowest loss of training effects in IQ was observed for the Milwaukee (2.75 IQ points per year) and Abecedarian (0.28 IQ points per year) projects respectively. Probably the inclusion of children with higher IQ scores facilitated the long-term training effects found for the Abecedarian project. For instance, studies considering a wide range of IQ scores have shown that children with higher IQ scores obtain more benefits from programs devoted to arithmetic (Hativa, 1988) and reading (Atkinson, 1974) skills, along with general academic performance (Cook et al., 1975; *Sesame Street'* program).

An important study was implemented in Venezuela (Herrnstein, Nickerson, de Sanchez, & Swets, 1986). 900 youngsters in seventh grade in a poor district of a Venezuelan provincial city were randomly divided into experimental and control groups (Herrnstein & Murray, 1994). Those in the experimental group were taught approximately sixty lessons for forty-five-minutes each in addition to their regular curriculum during the year. The training group received special enriched instruction in such cognitive skills such as classification learning, sequence evaluation, the analysis of verbal arguments, and a number of other cognitive skills (Hunt, 2012a). The training and control groups were tested in intelligence at the same intervals (before and after the year of training program). The difference between groups was of 1.6 IQ points in the Cattell Culture Free test (Cattell, 1940) and 6.5 IQ points in targeted abilities (Herrnstein & Murray, 1994). However, changes in the Venezuelan government made any follow-up impossible (Hunt, 2012a).

On the basis of the above findings, the APA accepted that improvements in intelligence are possible, but also that gains fade when the program is over (Neisser et al., 1996; Nisbett et al., 2012). In recent years (from 2008 to present), the number of studies analyzing the effect of cognitive training over intelligence scores has increased exponentially after the Jaeggi et al. (2008) report. Moreover, some authors consider the development of programs to improve and preserve people's intelligence a requirement for cognitive researchers in the early 21st century (Hunt & Jaeggi, 2013).

2. 5. New age programs

The new age programs can be classified into (a) pharmacological/nutritional, (b) educational and (c) psychological (Hunt & Jaeggi, 2013). Positive effects have been found for each of these types of intervention. The main difference with the pioneer programs is that the new programs were carried out with children (not only with children at high risk for mental retardation) and also with adults.

2.5.1. *Pharmacological/nutritional approaches.*

The effects of nutrition on intelligence have been widely reported (Hunt, 2012a). Nutritional deficiencies can have adverse effects on cognition after long periods of time without an appropriate diet (Neisser et al., 1996; Nisbett et al., 2012). Factors such as

reduction in caloric intake, protein deficits and parasitic and infectious diseases, have been shown to negatively affect cognition (Neisser et al., 1996; Nisbett et al., 2012). Evidence from developing countries suggests that nutritional supplements and public health policies that protect against disease can improve cognition (Hunt, 2012b; Schoenthaler, Doraz & Wakefield, 1986). In developed nations, it is possible to easily access appropriate nutrients, so the nutritional effects on cognitive development are mediated by choice of diet (Hunt & Jaeggi, 2013). For example, the Mediterranean diet has been suggested to reduce the risk of Alzheimer's disease (Solfrizzi et al., 2011). Similarly, omega-3 fatty acid supplementation can improve various aspects of cognition (Luchtman & Song, 2013).

Pharmacologically active substances and nutritional supplements have been used to raise intelligence. Nutritional supplements are, in general, more effective in preventing lowered cognitive performance than in enhancing normal or above normal performance (Eysenck, 1995; see Pollitt, Gorman, Engle, Martorell, & Rivera, 1993). There are some studies showing increments in intelligence for healthy people. For instance, Schoenthaler et al. (Schoenthaler, Amos, Eysenck, Peritz, & Yudkin, 1991; Schoenthaler, Bier, Young, Nichols, & Jansenns, 2000) administered vitamin/mineral supplements to a group of children while a placebo was given to a control group. The experimental group showed greater test score gains. Eysenck (1995) summarized the main results of these studies and concluded that 3.5 IQ points in fluid intelligence tests was the gain in the experimental group compared to the control group. But, the effects on people with deficits were even higher (10.5 fluid IQ points improvement).

Pharmacologically active substances that improve cognition or intelligence (nootropic) are still under study (Giurgen, 1973, Smith & Farah, 2011). The cognitive effects of pharmacological substances on normal healthy people cannot yet be characterized definitively, despite the volume of research that has been carried out (see Smith & Farah, 2011). Results have shown that declarative memory can be improved with some evidence consistent with enhanced consolidation of memories, but effects on working memory and cognitive control have been less reliable (Elliot et al., 1997; Kimberg, D'Esposito, & Farah, 1997; McGaugh & Roozendaal, 2009).

The number of studies showing an increment in intelligence test scores after a pharmacological intervention is less large. Rae, Digney, McEwan, and Bates (2003)

used a double blind, placebo-controlled, crossover design to study the increment in intelligence after applying an oral creatine supplement. This substance is expected to provide additional energy to the brain to improve the resources needed for a heavy cognitive workload. A significant improvement in intelligence scores after the administration period was found. However, this study is not generalizable, since the creatine was only applied to vegan/vegetarian participants and creatine levels are lower in these people than in omnivores. Recently, Stough et al., (2011) showed an increment of 6 IQ points in the Raven test after providing a nootropic agent in a double blind, placebo-controlled study. The pharmacological substance administered was Ceretrophin™, which is a combination of Huperzine, Vinpocetine, Alpha lipoic acid, Acetyl-L-carnitine, and *Rhodiola Rosea*. Other substances such as *Ritalin*® have been provided with the same results (for a review see, Lynch, Palmer, & Gall, 2011), although this drug is designed for the treatment of attention deficit disorder (ADD) and attention deficit hyperactivity disorder (ADHD).

These physical agents may have their effects due to their action on the brain. However, there are no long-term follow-ups. Nevertheless, pharmacology is a critical strategy for understanding the biological basis of intelligence, since it offers the opportunity to improve our knowledge regarding the cellular and neuronal mechanisms of intelligence differences (Stough et al., 2011). Lastly, the study of these substances is highly relevant as there is an increasing use of nootropic in western societies (Sahakian & Morein-Zamir, 2007; see Smith & Farah, 2011).

2.5.2. *Educational approach.*

Intelligence test scores and educational attainment are correlated (Herrnstein & Murray, 1994; Jensen, 1998; Neisser et al., 1996; Nisbett et al., 2012). School attendance may function as both a dependent and independent variable in relation to intelligence (Neisser et al., 1996). The main goal of education is the acquisition of specific pieces of knowledge and the acquisition of better reasoning capacities, which are highly relevant in our modern societies (Hunt & Jaeggi, 2013; Neisser et al., 1996). It has been suggested that education has a positive causal effect on cognitive ability (Ceci, 1991). Green, Hoffman, Morse, Hayes, and Morgan (1964) found a drop of 6 IQ points per year of education loss. Brinch and Galloway (2011) took benefit of the natural experiment created in Norway when an extra two years of schooling beyond the seventh

grade began to be required. Effects on IQ were substantial at the age of 19. However, although the relationship between intelligence and educational attainment has been documented, the nature of the relationship is elusive (Dolan et al., 2006). Indeed, a recent report showed that education does not improve general intelligence (*g*), but rather specific cognitive skills assessed by the tests (Ritchie, Bates, & Deary, 2015).

There are different educative programs. For example, the pioneer programs commented above could be considered a mix between educative and nutritional programs. The effects of these programs have been amply discussed (e.g. Ceci, 1991; Herrnstein & Murray, 1994). Some authors suggest that their benefits are clear (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002), but the discrepancy between the effects found for school achievement and IQ indicates that observed changes could be attributed to improvements in self-control, attention, and perseverance rather than to intellectual gains per se (Knudsen, Heckman, Cameron, & Shonkoff, 2006). De Acedo Lizarraga, Ugarte, Iriarte, & Sanz de Acedo Baquedano (2003) implemented an extensive educational intervention using the materials developed in the context of Venezuelan government project (Herrnstein et al., 1986). They found a significant increment of scores in several measures of intelligence. However, they included a variety of tasks intended to enhance self-regulation, and, therefore, it is uncertain which aspects of their intervention increased intelligence scores (Nisbett et al., 2012). Another example of such educational programs is "*Philosophy for Children*", which is aimed to develop higher order thinking skills. This program was developed by Matthew Lipman during the 70's, and has been implemented in different countries. Meta-analysis carried out on this program revealed a positive effect (d = .58) over different tests. However, these effects were higher in tests designed to assess the program (d = .63) than in composite intelligence scores (d = .31), but, nevetheless, the effects were relevant and statistically significant in both cases (García-Moriyón, Rebollo & Colom, 2005).

Finally, some programs were based on test-specific training. These coaching programs were mainly focused in the improvement of different standardized tests administered in the admission processes of USA colleges. SAT (The Scholastic Achievement Test) is the most employed test. Messick and Jungeblut (1981) reviewed the published studies on coaching for the SAT. They estimated the effect of spending up to 300 hours on either a verbal or a math test in a coaching course (including homework). They

observed that 30 hours of studying for either test led an average improvement of sixteen points on the verbal SAT and twenty-five points for the math SAT, whereas 100 hours led to an average improvement of twenty-four points on the verbal SAT and thirty-nine points on the math SAT. Even 300 hours-two additional hours for 150 days-can be expected to lead to gains of only seventy additional points on the combined score. Herrnstein and Murray (1994) concluded that these results were similar to those obtained in the study of Venezuela's government, since the 45 hours of this study suggested an increment of 1.5 to 6 IQ points. The same amount of time led to an increment of 2.4 to 3.45 IQ points in SAT (Verbal and Math respectively). However, Atkinson (2005) suggested that reported increments in the SAT were test-specific (Brown & Campione, 1982) and could not be considered as a true increment in intelligence.

In conclusion, educational programs are based on learning/training of different skills during long periods of time. To detect a positive effect caused by the program, it is required to assess the transfer to a wide range of cognitive abilities (Buschkuehl & Jaeggi, 2010; Jensen, 1998; Hunt, 2011). Furthermore, the training tools must be substantially different to usual tests of cognitive ability, if one aims for convincing evidence, since skills can simply be improved by increased familiarity rather than by a genuine improvement in intelligence (te Nijenhuis et al., 2007).

2.5.3. *Psychological approach.*

The number of studies aimed at testing whether intelligence can be raised through cognitive training programs has grown in the last seven years (Au et al., 2014; Buschkuehl, & Jaeggi, 2010; Colom et al., 2010a; Estrada et al., 2015). There are three main differences between educational and psychological approaches: (a) the number of training and practice sessions are generally greater in the former, (b) the educational approach implies an investment of a lot of time in the acquisition of specific knowledge and skills, while the psychological approach focuses on basic cognitive process, and (c) the psychological approach relies on computerized programs.

Why has the number of studies based in training cognitive processes been increasing in the last years? Sternberg (2008) pointed out that advances in cognitive theories of intelligence did provide insights into what kind of training might be successful to

promote intelligence. Indeed, most of the studies have been based on working memory training, mainly because of its very high correlation with Gf (Ackerman et al., 2005; Colom et al., 2004, 2005, 2008; Cowan et al., 2005; Hunt, 2011; Kyllonen & Christal, 1990; Martínez et al., 2011; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Nevertheless, there are programs aimed at different cognitive process, such as attention control or processing speed, or even based on neuro-feedback methods (see Keizer, Verschoor, Verment, & Hommel, 2010).

2.5.3.1. Cognitive programs: Attention and processing speed programs.

Tang and Posner (2009) suggested that computer-based exercises might lead to improved attention in children and adults. But are these improvements related to increased intelligence levels? Posner and Rothbart (2007) trained children on a visual attention task based on the management of conflict. Their main outcome was that trained children obtained a higher score in a standard intelligence battery than controls. Moreover, Rueda, Rothbart, McCandliss, Saccomanno, and Posner (2005) coached two groups of children (4 and 6 year-olds) in different adaptive attention tasks. They obtained an increment in intelligence scores for the training group in comparison with an untrained control group. However, this result was found only in the 4 year-olds group. Rueda and colleagues argued that this effect was produced because attention tasks and intelligence tests share the same neural network and explained their findings on the basis of brain plasticity being higher in youngest children. So, far-transfer effects are expected to be more difficult with increasing age (Garlick, 2002). However, Karbach and Kray (2009) investigated this issue of task-switching in children, young adults, and old adults. All groups obtained significant increments in intelligence performance.

Colom et al., (2010a) coached a group of 115 young adults across eighteen practice sessions (6 x 3) with simple processing speed and attention control tasks, finding germane increments in intelligence score performance after practice. However, these increments were found to be hollow in terms of the general factor tapped by the intelligence tests. Indeed, analyzing this dataset using a latent-variable approach, Estrada et al., (2015) found that changes at the test level were not mediated by changes in the general high-order intelligence factor (*g*). Similar results were observed for

practice sessions with highly complex short-term and working memory tasks (Colom et al., 2010a; Estrada et al., 2015).

2.5.3.2. Cognitive programs: Working memory programs.

It has been shown that working memory capacity (WMC) can be trained (Diamond & Lee, 2011; Klingberg, 2010; Morrison & Chein, 2011). Working memory tasks involve the preservation of information in the short-term while simultaneously performing distracting or interfering tasks (Case, Kurland, & Goldberg, 1982). In this regard, working memory capacity taps the limit of an individual's ability to save a given amount of information in the presence of interference (for a review, see Conway et al. 2005). Kyllonen and Christal (1990) proposed that WMC is responsible for differences in reasoning ability, because the processes involved in working memory are central to the processing stages of reasoning tasks. However, their results may be interpreted as supporting the hypothesis that WMC is mainly determined by individual differences in fluid intelligence. Colom et al., (2004) analyzed three datasets and found that WMC is the latent factor best predicted by *g*.

The seminal study by Jaeggi et al., (2008) was based on the hypothesis that working memory and fluid intelligence share overlapping cognitive processes (Halford, Cowan, & Andrews, 2007). The shared capacity limitations are related to the amount of items that can be kept active in working memory or the number of relationships between elements that can be combined during the reasoning process necessary for solving the problems found in intelligence tests. These limitations could be based on the ability to build and keep binding among items in the short-term (Colom et al., 2004, 2005, 2008; Kyllonen & Christal, 1990; see also Carpenter, Just, & Shell, 1990).

It has been considered that attention control processes may drive this common capacity limit (Halford et al., 2007; Jaeggi et al., 2008, 2010a). However, the exhaustive analysis performed by Colom et al., (2008), based on the simultaneous consideration of short-term storage, mental speed, updating, and the control of attention along with working memory and intelligence measures across three separate studies, revealed that the high relationship between WMC and intelligence was essentially mediated by short-term memory (STM) (see also Martínez et al., 2011).

On the other hand, there is evidence showing that WMC and intelligence share neural networks (e.g. Barbey, Colom, Paul, & Grafman 2014a; Colom et al., 2007; Gläscher et al., 2010). This supports the presumption that practice on tasks requiring working memory capacity would stimulate neural circuits involved in intelligent behavior (Hunt, 2012a). This particular issue will be discussed in detail later in this report.

It is important to highlight that not all working memory programs are equal. There are several criteria that should be met:

(1) Task must minimize the use of domain-specific strategies (Buschkuehl & Jaeggi, 2010; Jaeggi et al., 2010a; Klingberg, 2010; Morrison & Chein, 2011).

(2) Programs must minimize automaticity, meaning that training must be adaptive to allow participants to reach and train at the peak of their performance (Buschkuehl & Jaeggi, 2010; Jaeggi et al., 2010a; Morrison & Chein, 2011).

(3) Programs must be based in WMC tasks (Klingberg, 2010), which include multiple modalities (e.g. verbal, numerical, spatial) in order to maximize process overlap with other tasks (Buschkuehl & Jaeggi, 2010; Jaeggi et al., 2010a; Morrison & Chein, 2011).

(4) High cognitive workloads and high intensity cognitive engagement is required (Jaeggi et al., 2010a; Morrison & Chein, 2011).

(5) Training schedules should be rigorous and include roughly at least 20 sessions, lasting 30–60 min each (Klingberg, 2010).

The dual n-back task used by Jaeggi et al., (2008) fulfills all these criteria. However, the number of tasks employed in working memory programs is very wide (Morrison & Chein, 2011; Shipstead, Redick, & Engle, 2012). Examples of tasks employed in these kinds of programs are shown in Figure 2.3.

As observed, even when all these tasks measure WMC, several differences are evident. Therefore, far-transfer to constructs such as fluid intelligence may differ as a function of the tasks used. For example, while some working memory tasks comprise active and strategic recall processes, others rely on passive recognition processes (Jaeggi et al., 2010a). Indeed, available evidence provides disparate results: (a) positive transfer to Gf measures (e.g., Irwing, Hamza, Khaleefa, & Lynn, 2008; Jaušovec & Jaušovec, 2012;

Klingberg, Forssberg, & Westerberg, 2002; von Bastian & Oberauer, 2013) and (b) lack of transfer to Gf measures (e.g., Holmes, Gathercole, & Dunning, 2009; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg 2009). Reviews support both conclusions (Buschkuehl & Jaeggi, 2010; Dahlin, Bäckman, Neely, & Nyberg, 2009; Diamond & Lee, 2011; Klingberg, 2010; Morrison & Chein, 2011; Perrig, Hollenstein, & Oelhafen, 2009; Shipstead et al., 2012).

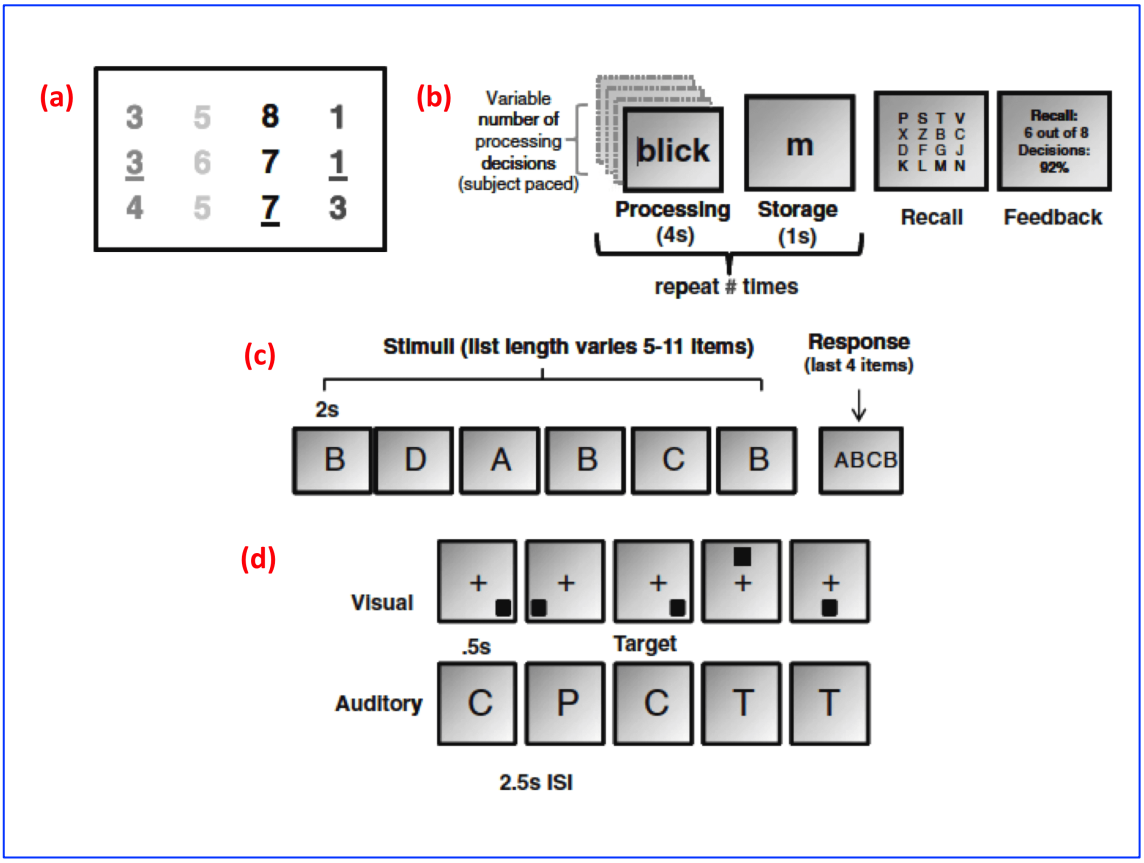


Figure 2.3. Examples of task employed in different working memory training programs. (a) Schematic drawing of the identity-judgment N-back task. In this task, participants are asked to determine if each sequentially presented digit is the same as the digit shown N items back. (b) Schematic drawing of the verbal condition of the complex WM span task. Here, participants are asked to remember the items in the storage condition while making intermittent processing judgments; the stored items are reported at the end of the trial. (c) Schematic drawing of the letter-updating task. In this case, several lists of items with an unknown length are presented to participants and they are asked to report the last four items. (d) Schematic drawing of 2-back condition of the dual N-back task, which requires that participants simultaneously complete auditory and visual iterations of an N-back task. (Figure adapted from Morrison & Chein, 2011).

Melby-Lervåg and Hulme (2013) carried out a meta-analysis on the effect of working memory programs and their transfer to intelligence measures. They analyzed only the studies that followed the criteria recommended by Klingberg (2010) (see above). In the

meta-analysis, they focused on two types of transfer: (a) near transfer (for example, improvements on a visual-spatial working memory task after practicing on a verbal working memory task) and (b) far transfer effects on tasks quite different from those trained (e.g., improvements on IQ tests following training on working memory tasks). The main findings were: (a) a large transfer effect (d = 0.79) to verbal working memory tasks (near-transfer), (b) a moderate effect (d = 0.52) to visual-spatial working memory tasks (near-transfer), (c) small effects to nonverbal abilities (d = 0.19), verbal abilities (d = 0.13), word decoding (d = 0.13) and arithmetic (d = 0.07) (far-transfer), and (d) a moderate effect (d = 0.32) to inhibitory processes in attention measured by the Stroop task. Therefore, they concluded that working memory programs fail to improve performance on far-transfer measures, although they might improve performance on near-transfer measures.

Here, we are interested in the n-back training programs carried out with healthy young adults. Therefore, general reviews are not particularly interesting. The next section focuses on the available evidence in this respect.

2.5.3.3. Cognitive programs: N-back programs with young people.

The report by Jaeggi et al. (2008) opened a new age for studies on working memory training and their transfer to fluid reasoning. This study showed that training with the dual n-back task (Figure 2.4) improves scores on fluid intelligence. The Dual n-back task recruits multiple processes, such as monitoring, encoding of the incoming stimuli, updating and maintenance. Also, selection, inhibition, decision and interference resolution processes are involved (Jonides et al., 1997). It has been argued that the n-back task requires the retrieval of items that are no longer in the focus of attention, thus it requires a shift of attention (McElree, 2001; Verhaeghen & Basak, 2005; Verhaeghen, Cerella, & Basak, 2004). It has been proposed that dual tasks are good estimates of WMC because they preclude the use of strategies (Oberauer, Lange, & Engle, 2004).

The n-back task was initially introduced by Kirchner (1958) as a visual-spatial task with four load factors (‘‘0-back’’ to ‘‘3-back’’), and by Mackworth (1959) as a visual letter task with up to six load factors. This task is one of the most popular WMC measures in neuroimaging (Conway et al., 2005; Jaeggi et al., 2010b).

Jaeggi et al. (2008) trained 70 healthy young adults divided in four groups depending on the duration of the training program: 8 days (N = 16), 12 days (N = 22), 17 days (N = 16), and 19 days (N = 15). Improvements attributable to the training are depicted in Figure 2.5 (top panel). The transfer to fluid intelligence (far-transfer) was assessed by the Raven or BOMAT tests (Figure 2.5a). Near-transfer was also evaluated using the digit-span and reading-span tasks. Importantly, improvements on Gf scores were susceptible to a dose effect (Figure 2.5b).

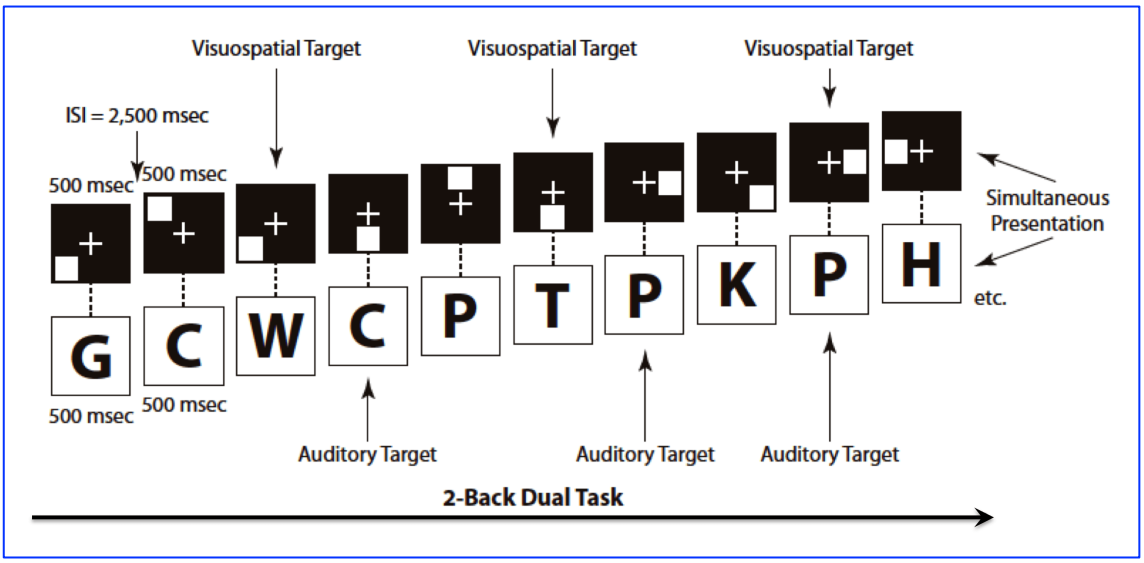


Figure 2.4. Example of dual n-back task (2-back). A stream of sequential stimuli are presented to participants. When the stimulus in the current screen is the same as the one presented n stimuli back in the stream, the participant is required to press a key. A visual and an auditory modality are presented simultaneously (Figure adapted from Jaeggi, Buschkuehl, Perrig, & Meier, 2010b).

The reactions to this study were disparate (e.g. Haier, 2014; Melby-Lervåg & Hulme, 2013; Moody, 2009; Shipstead et al., 2012). These were the main criticisms: (a) the differences in Gf were circumscribed to a single test (BOMAT), (b) the administration time of Gf tests was very short (10 min vs. 45 min in the standard administration time), and, therefore, participants were not allowed to face the most difficult items, (c) the control group was passive (no-contact), and (d) the near transfer effects were lacking.

Several studies were carried out to replicate the main findings of the Jaeggi et al (2008) study but the results were mixed:

(a) No transfer found. Chooi and Thompson (2012) used several measures for estimating changes after training on the dual n-back task modeled by Jaeggi et al. (2008). They failed to find any effect of training on either intelligence or working memory (far and near transfer). Redick et al. (2012) reported a similar study: fluid intelligence (six tests) and crystallized intelligence (two tests), along with working memory (two tasks), multitasking (three tasks), and processing speed (two tasks) were the constructs measured. Again, no significant effects were found. Heinzel et al. (2014) administered a battery of tests/tasks (short term memory, episodic memory, processing speed, executive functions, and fluid intelligence), finding transfer only to short-term memory (Digit symbol) and executive function tasks (see also Oelhafen et al., 2013; Salminen, Strobach, & Schubert, 2012; Thompson et al., 2013).

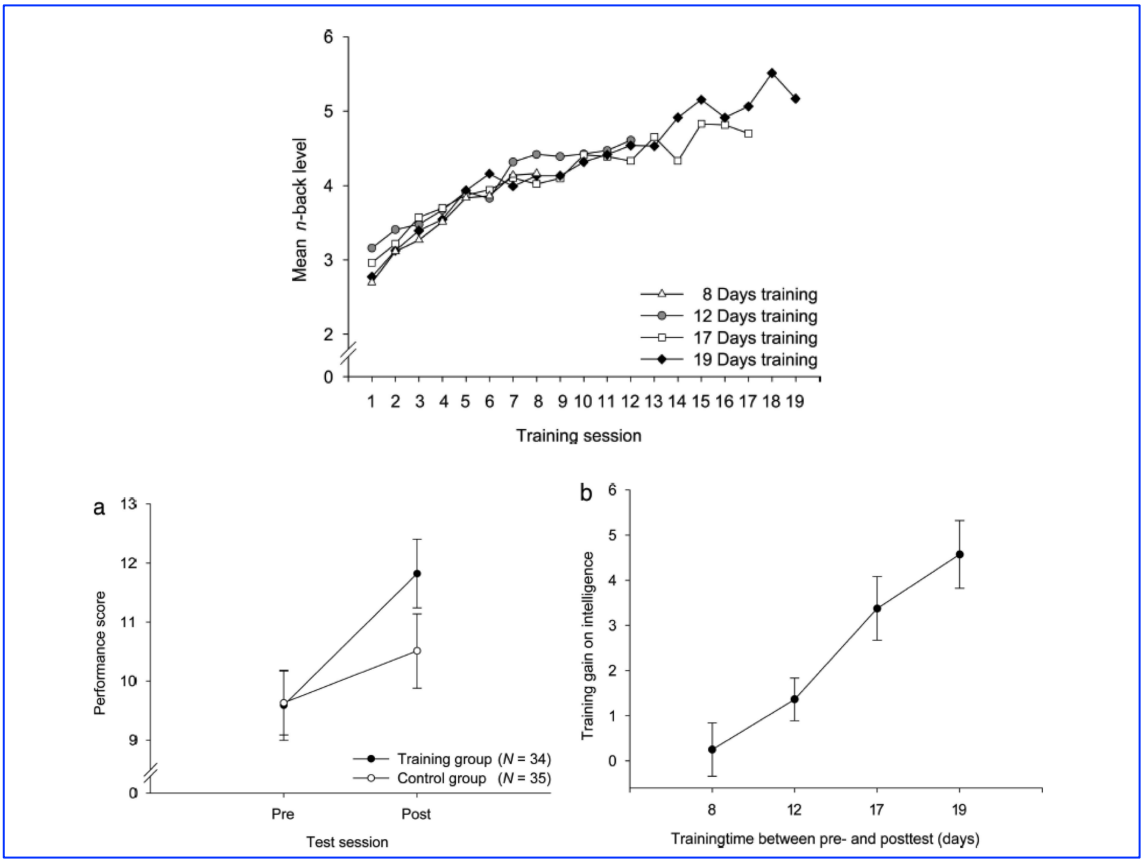


Figure 2.5. Top panel shows n-back training performance. Bottom panel (a) increment in Gf for the training and control groups, and (b) Gf gains related with number of training sessions (Jaeggi et al., 2008).

(b) Transfer found. Jaeggi et al., (2010a) found significant far-transfer to Gf measures after training based on dual and single n-back tasks, although near-transfer to complex working memory tasks was not found. Jaeggi, Buschkuehl, Shah, and Jonides (2014)

found transfer to a composite measure of visual-spatial reasoning, even after controlling for motivation, need for cognition, preexisting ability, and implicit theories about intelligence. However, this transfer effect was not found for verbal reasoning tests. They argued that people generally have less practice on spatial than on verbal tasks, and thus they may have more room for improvement on spatial tasks. Other laboratories found similar results. For instance, Jaušovec and Jaušovec (2012) reported a positive effect in the Raven test equivalent to thirteen IQ points (d = .88) for the training group, whereas it was null for an active control group. Digit span scores were also substantially higher for the training group (d = 0.81) than for the control group (d = 0.25). Rudebeck, Bor, Ormond, O'Reilly, & Lee, (2012) concluded that training in dual n-back task is enough to find effects on fluid intelligence and episodic memory. Therefore, an extensive training regime with multiple tasks may not be necessary. Further studies reported similar transfer effects (e.g., Kundu, Sutterer, Emrich, & Postle, 2013; Schweizer, Hampshire, & Dalgleish, 2011; Stephenson & Halpern, 2013; Vartanian et al., 2013). Finally, even when motivation effects seem crucial for a satisfactory training, Katz, Jaeggi, Buschkuehl, Stegman, and Shah (2014) found that the inclusion of motivational features during n-back training was not beneficial.

An analysis of these studies shows that (a) transfer is not circumscribed to the BOMAT test, since short-term memory, executive functions and even the Raven test have shown significant interaction (training vs. control) after n-back training (Heinzel et al., 2014; Jaeggi et al., 2010a; Jaušovec & Jaušovec, 2012; Kundu et al., 2013; Salminen et al., 2012; Stephenson & Halpern, 2013), (b) passive control groups show equivalent changes to those observed for active control groups (Chooi & Thompson, 2012; Redick et al., 2012), and (c) the lack of transfer to working memory complex task might be due to a rather low correlation between n-back and other WMC tasks (Jaeggi et al., 2010a,b; Kane, Conway, Miura, & Colflesh, 2007; Schmiedek, Hildebrandt, Lövdén, Lindenberger, & Wilhelm, 2009).

A recent meta-analysis by Au et al. (2014) focused on studies using the n-back task, has demonstrated a small positive impact on fluid intelligence. The weighted average effect size was d = .24 (3.6 IQ points). But as noted above (section 2.2), even these small increments might impact performance in real-life settings (Herrnstein & Murray, 1994):

"*Since Gf is a fundamental cognitive skill that underlies a wide range of life functions, even small improvements can have profound societal ramifications, particularly given the healthy young adults in our analyses, representative of society's workforce. Taken together, it is becoming very clear to us that training on WM with the goal of trying to increase Gf holds much promise*" (Au et al., 2014).

Their analysis of the sources of variance across studies revealed that (a) International studies tended to find more transfer than U.S. studies, (b) there was no difference between the research conducted by the Jaeggi's team and other groups, (c) the type of control group used in studies (active or passive) did not moderate the training effect, and (d) the hypothesis that money compensation affects negatively to training effect was not supported after the elimination of outliers. However, this meta-analysis did not study the maintenance of these effects. This issue is discussed in the conclusion section.

2. 6. Conclusion.

Intelligence can improve (Hunt, 2012a; Hunt & Jaeggi, 2013). Here we have focused on intervention programs aimed at raising intelligence. The pioneering programs were devoted to high-risk children (Head Start, Milwaukee, Abecedarian), but their results were highly arguable. New interventions based on different approaches (pharmacological/nutritional, educational and psychological) have emerged in recent years. The pharmacological/nutritional approach is aimed at improving the central nervous system, the educational approach trains cognitive skills, and the psychological approach is focused on improving basic cognitive processes (Hunt, 2012a). Satisfactory transfer effects validate all these approaches over several intelligence tests.

After Jaeggi et al.'s (2008) study, the number of reports and critical analyses has grown quickly (Buschkuehl & Jaeggi, 2010; Conway & Getz, 2010; Haier, 2014; Moody, 2009; Shipstead, et al., 2012). It is presumed that the effect of these training programs may help to prevent cognitive decline. Recent studies showed that the older people (even with mild cognitive impairment or Alzheimer's disease) could improve their scores in different domains as: visual-spatial memory, audio-verbal memory, executive function, and verbal fluency (Fernandez-Ballesteros et al., 2003, 2012). These authors found that there was a gradient of modifiability from healthy to mild cognitive impairment and from mild cognitive impairment to Alzheimer's disease, but even

Alzheimer's disease patients benefited from training. Other laboratories have employed working memory tasks with the goal of raising intelligence scores. For example, Schmiedek, Lövdén, and Lindenberger (2010) found transfer effects to RAVEN test, although the changes in Gf latent factor were not statistically significant. The literature on this topic is wide and exceeds the scope of this chapter. However, as longevity rises, the rate of severe loss of intelligence associated with old age poses a main problem in public health (Hunt & Jaeggi, 2013).

Several commercial, computer-based and working memory training programs have been developed (e.g. *CogMed*: http://www.cogmed.com/, *Jungle Memory*: http://www.junglememory.com/, *mindsparke*: http://www.mindsparke.com, *lumosity*: http://www.lumosity.com, *logisera*: http://www.logisera.com), although scientific proof supporting them is rarely provided (Stanford Center on Longevity, 2009). Nevertheless, studies with commercial videogames showed that performance on some videogames requires intelligence (Quiroga et al., 2009, 2011, 2015). Also, it has been shown that practice on videogames is related to structural and functional brain changes (Colom et al., 2012; Martínez et al., 2013; Kühn, Gleich, Lorenz, Lindenberger & Gallinat, 2014; Nikolaidis, Voss, Lee, Vo, & Kramer, 2014).

Hence, this kind of intervention should be considered similarly to cardiovascular training (Jaeggi et al., 2014). Consequently, as in physical training, the effects would remain if and only if the training program is maintained (Vallejo-Nájera & Colom, 2004). The same applies for pharmaceutical and dietary enhancements (Hunt, 2012a). Therefore, practice or booster sessions are required for maximizing retention, as in physical training (Bell et al., 2008; Haskell et al., 2007). Indeed, the follow-up studies allow checking the duration of benefits of cognitive training programs. The meta-analysis conducted by Melby-Lervåg and Hulme (2013) with all WMC programs showed that: (a) long-term effects on verbal working memory are small after removing outliers (d = 0.10), (b) medium (d = 0.41) on spatial working memory and (c) small far-transfer effects can be detected for nonverbal ability (d = -0.06), attention (d = 0.09), decoding (d = 0.13) and arithmetic (d = 0.18). Their main conclusion was that the effects found after working memory training do not persist at follow-ups.

Regarding n-back training programs, the effects were mixed. Studying children, Jaeggi, Buschkuehl, Jonides and Shah (2011) found that the effect on Gf measures remained

intact after 3 months. However, Jaeggi et al. (2014) failed to replicate this result with young adults. Similar outcomes were found by Thompson et al. (2013), since improvements on the trained tasks persisted for at least 6 months after training but a transfer was not observed to any of the non-trained measurements when compared to a third untrained group serving as passive controls. In contrast, Li et al. (2008) showed conservation of practice gains and of near-transfer effects at 3-month follow-up in young adults. However, older participants showed a performance decrement from post practice to follow-up. Therefore, the evidence is far from clear.

Research would reveal ways of keeping the main mechanisms supporting trait intelligence and everyday life performance (Willis & Schaie, 2009). As noted by Hunt and Jaeggi (2013):

"*Arguments over whether or not intelligence is malleable are obsolete. […] there is every reason to be optimistic about future efforts to improve and/or maintain intelligence, in the general sense of individual intellectual competence. Improving intelligence, in the much narrower sense of increasing intelligence test scores, is a side issue* " (Hunt & Jaeggi, 2014, p. 49).

CHAPTER 3: Neuroimaging Framework for Studying Brain Plasticity.

3. 1. Introduction.

"*Plasticity is an intrinsic property of the human brain and represents evolution's invention to enable the nervous system to escape the restrictions of its own genome and thus adapt to environmental pressures, physiologic changes, and experiences*" (Pascual-Leone, Amedi, Fregni, & Merabet, 2005).

William James (1890) was the first introducing the notion of neuronal plasticity. Later, Santiago Ramón y Cajal (1904) proposed that behavioral changes have an anatomical basis in the brain and extended the notion of plasticity to the neural substrate.

The scientific community currently accepts that plasticity is an intrinsic property of the nervous system. In fact, plasticity is the standard state of the nervous system throughout the life span (Pascual-Leone et al., 2005).

Changes in the brain can be due to: (a) typical developmental sequence, which can show individual variations (Dennis & Thompson, 2013); (b) experience, because changes at behavioral level reflect changes in the brain (Gazzaniga, 2004); and (c) direct intervention of physical agents, which may either be beneficial or deleterious (Hunt & Jaeggi, 2013). These 'changes in the brain' are manifested with the growth/death of neurons, or modifications in the paths connecting single neighboring neurons or long distance groups of neurons.

This work studies the changes at behavioral and biological levels after completing a cognitive training program based on the n-back task (Jaeggi et al., 2008). Changes at a biological level will be quantified by changes in T1 and Diffusion-Weighted images obtained by Magnetic Resonance imaging (MRI), since it has been suggested that changes in the MRI signal may reflect changes in axonal myelination, neurogenesis, angiogenesis, dendritic spine motility, glial cell proliferation, and synaptogenesis (Draganski & May, 2008; Scholz, Klein, Behrens, & Johansen-Berg 2009), which reflect brain plasticity. Specifically two processing approaches will be used: (1) surface-based morphometry (SBM) for obtaining two grey matter metrics (cortical thickness

and cortical surface area) and probabilistic tractography for building white matter structural connectivity (SC) matrices.

This chapter introduces basic notions about the acquisition of MR images, as well as the processing protocols required to obtain cortical thickness, cortical surface area and SC matrices. Next, available evidence regarding brain changes due to lifespan development and after diverse training programs is summarized. Finally the brain networks shared by intelligence and working memory is discussed, because this work is also aimed at exploring the transfer of n-back training effects to intelligence.

3. 2. Magnetic Resonance Imaging (MRI): Acquisition

In this section, the main steps for acquiring T1 and diffusion weighted magnetic resonance images are briefly described (see Hashemi, Bradley & Lisanti, 2004; Lipton, 2008; Martínez, 2014 for further details). T1 and diffusion weighted images were used in this study to check for changes in grey and white matter respectively.

Brain MRI is a noninvasive technique based on the principles of nuclear magnetic resonance. The magnetic nuclei in a magnetic field absorb and re-emit electromagnetic radiation depending of the strength of the magnetic field and the magnetic properties of the isotope of the atoms.

There are different kinds of images that can be obtained from a MRI scanner, such as: (a) functional images, which are used for measuring the brain response to a given task or intrinsic spontaneous neuronal activation; (b) structural brain imaging such as T1 and T2-weighted imaging, which allows studying several brain structure properties; and (c) diffusion-weighted image, which can be processed to reconstruct white matter pathways. Brain plasticity can be observed with the three type of images, but our research questions are related with (b) and (c).

3. 2. 1. *Imaging Hardware*

The fundamental hardware components of magnetic resonance scanners are: (1) the magnet, (2) gradient coils, (3) radiofrequency (RF) transmission coils and (4) RF reception coils (Figure 3.1).

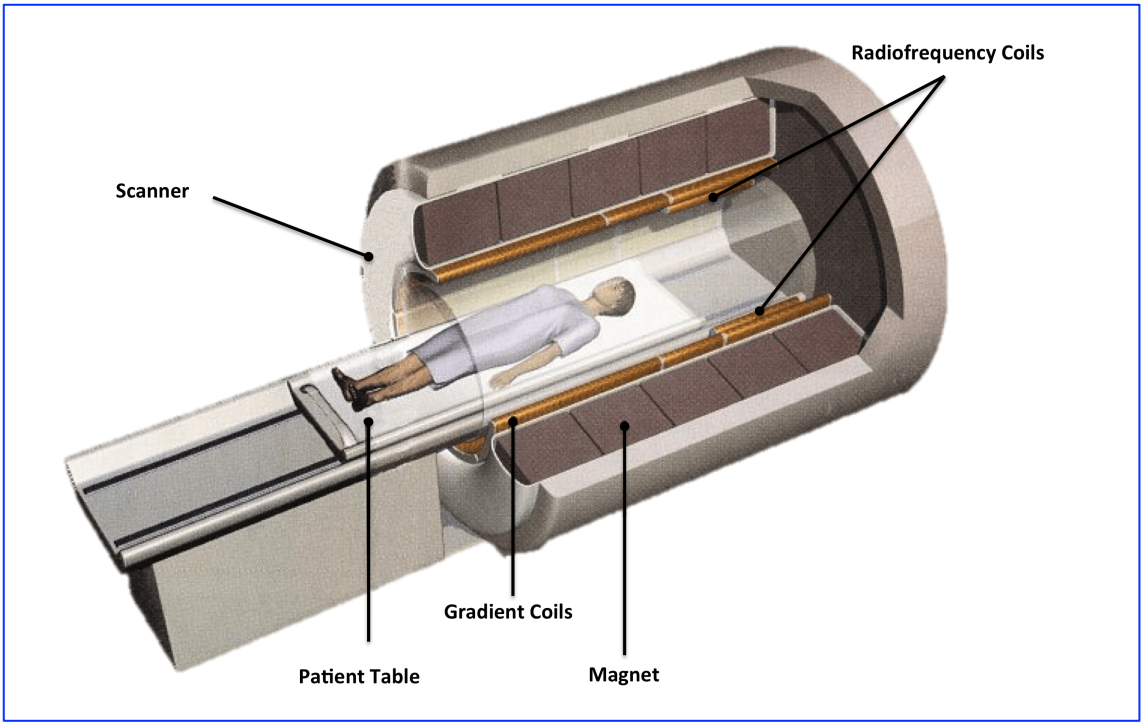


Figure 3.1. Components of MRI scanner adapted from http://miriam-english.org/files/IncredibleShrinkingScanner.html.

The magnet is usually a ring-shaped coil that generates a strong magnetic field (B_0). Its intensity is variable, although in neuroimaging the most common are 1.5 and 3 Tesla magnetic fields. Image resolution and acquisition time are related to the strength of the magnetic field. Usually, higher strength translates into faster acquisition and higher image resolution.

The gradient coils are three sets of orthogonally positioned magnets within the main magnet, so that each of them corresponds to an axis in three dimensions (x, y, z). Gradients are responsible for generating small magnetic fields (approximately 5% of the intensity of the main field) that determine the direction in which the image is acquired (for example, sagittal, axial, coronal, or any combination thereof). RF transmission coil are used to emit an excitation radiofrequency pulse (B_1). RF reception coils capture changes in the biological tissues after the elimination of B_1 field (see below for a broad description of the different steps).

3. 2. 2. *Basic MRI principles.*

MRI is based on the ability of certain particles in the human body (specifically, hydrogen atoms) to be excited by radiofrequency waves in the presence of a strong

magnetic field. The biological tissues are very rich in water (H_2O) and, therefore, the hydrogen nuclei ($_1H$) of the water molecules are the body's most abundant elements. The protons of the hydrogen atoms exhibit an inherent angular momentum (*spin*) around a longitudinal axis, yielding a magnetic moment (Figure 3.2a). Spinning hydrogen protons are like little magnets oriented randomly. With the lack of an external magnetic field, the spin directions of all atoms are random and cancel each other, resulting in a net magnetic moment of zero magnitude.

The first step to obtain an MRI image is to switch on a strong magnetic field (B_0) and, as a consequence; the spins align to B_0 field (equilibrium; Figure 3.2b)[1]. The sum of all tiny magnetic fields of each proton produce a net bulk magnetization (M) oriented in the same direction as the main magnetic field.

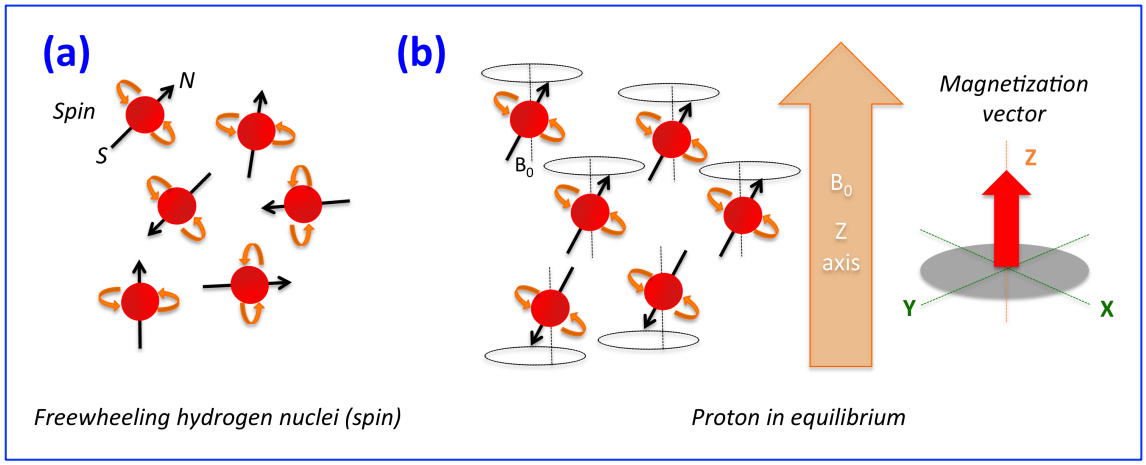


Figure 3.2. (a) Protons randomly aligned, (b) protons in equilibrium after putting in place a strong magnetic field (B_0) in Z-axis. Then, the spins align to B_0 field and the sum of all tiny magnetic fields of each proton produce a net bulk magnetization (M) in the same orientation than the strong magnetic field. N = north, S = South, B_0 = magnetic field.

The second step is that the RF transmission coil emits an excitation radiofrequency pulse (B_1) that deviates the nuclei from their equilibrium state by some degrees (flip angle) perpendicular to the longitudinal axis. During this process, the proton absorbs this energy (excitation).

Thirdly, the excitation RF pulse stops and the nuclei return to their original position under the influence of the magnetic field B_0. This returning to the initial position is

[1] The magnitude is usually of 1.5 or 3 Tesla. In this work, we obtained the images with a scanner of 3 Tesla

called longitudinal relaxation time. At the same time, a second phenomenon occurs. This is called transverse relaxation. It is produced by a progressive loss of magnetization due to dephasing. Relaxation results in a recovery of magnetization in the longitudinal plane and a decay of magnetization in the transverse plane. Therefore, hydrogen atoms start to be out of phase. The reason for this phenomenon is the interaction between proton magnetic neighbors.

Fourthly, while the nuclei are returning to their initial position, they emit energy that is captured by the receivers or antennas (RF reception coils). The combination of the signals received from all points of the brain is used for building the final image. With signals characteristic of certain types of brain tissue (e. g., white matter, grey matter, etc.), it is possible to generate contrast images within a grey scale, where differences remain bounded between tissues of interest. The time the proton needs to relax longitudinally is called T1, and varies depending on the type of tissue. The time needed for the proton transverse relaxation is called T2. It's important to fix the T1 and T2 to detect differences between tissues, since too long or too shorts T1s or T2s would not be good to detect differences between grey, white and cerebrospinal fluid.

Finally, when both processes are finished, a new excitation RF pulse is emitted and the process is repeated. This succession of steps is called *pulse sequence*, which contains radiofrequency (RF) pulses and gradient pulses with controlled durations and timings. The interval between the second step (RF pulse; B_1) and the reception of signal produced by protons in the receivers or antennas (fifth step) is called echo time (TE). The repetition time (TR) is the time between the emissions of two RF pulses (B_1) (Figure 3.3).

These times can be modified in order to get the required image contrasts. The choice of these times is highly relevant in obtaining images of high quality, since different tissues have different relaxation times. T1 images are especially important in the acquisition of images with high definition between structures, while T2 are especially useful for detecting pathologies (Figure 3.4).

T1-weighted images are highly essential to our study, since they are mandatory for the SBM protocol used to estimate cortical thickness (CT) and cortical surface area (CSA). Also, diffusion-weighted images (DWI) are used for the reconstruction of white matter

fibers. While T1 and T2 are obtained from signal relaxation times after excitation, DWI images are based on the translational motion of water molecules.

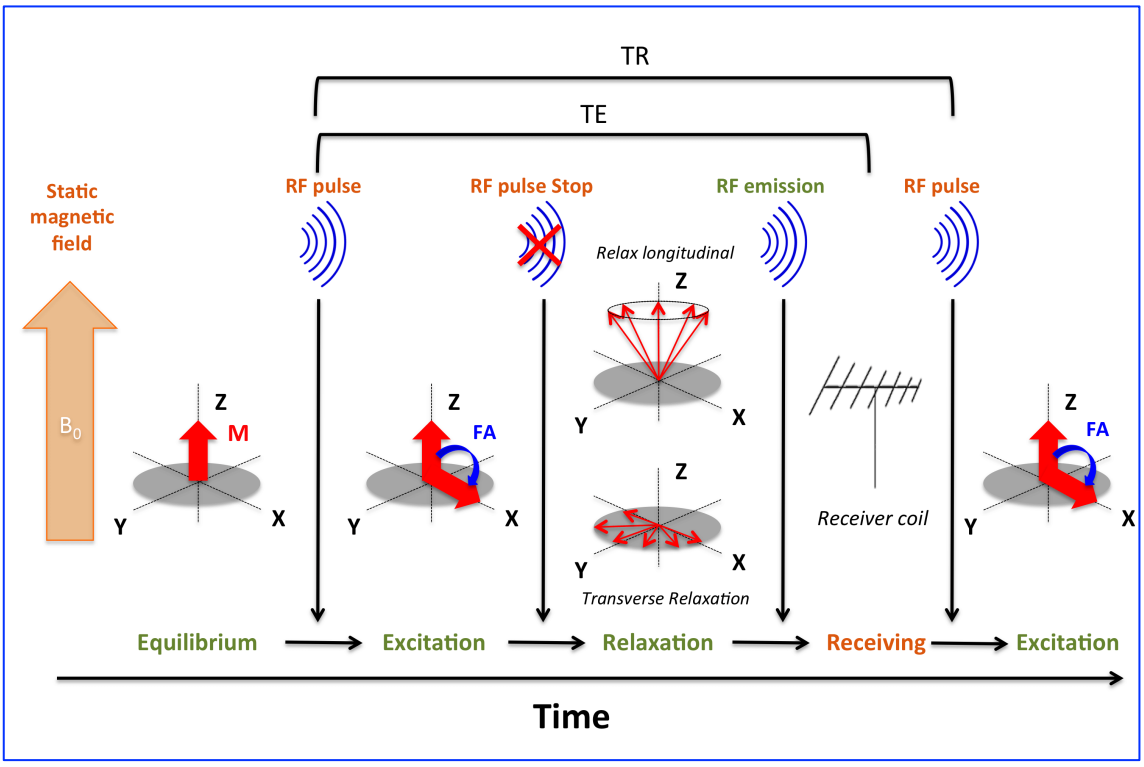


Figure 3.3. Pulse sequence of MRI acquisition. The steps are: (1) the subject is placed in a strong magnetic field and the spins align to B_0 field (equilibrium); (2) a excitation radio frequency pulse is emitted and atoms absorb energy (excitation): (3) as the excitation RF pulse stops, the relaxation processes begin; (4) the receiver coils 'listen' to atoms radio frequency emission due to relaxation; then (5) a new excitation RF pulse is emitted and the process is repeated. Steps in orange refer to actions that take place in the scanner (imaging hardware), whereas steps in green refer to different physic phenomenon that occurs due to the action of the imaging hardware. B_0 = magnetic field, RF = radiofrequency, M = net bulk magnetization, FA = flip angle, TE = echo time, TR = repetition time. Adapted from Martínez (2014).

DWI images are computed thanks to the application of a gradient pulse perpendicular to the longitudinal axis of B_0 when the protons are in phase (Figure 3.2b). The application of this gradient pulse causes a difference in frequency of precession of protons depending on their position in the longitudinal axis gradient (the closer the emitter, the higher the frequency increase). After the application of the gradient pulse, the protons return to the same frequency, but now out of phase (dephase). At this point, a second but reverse gradient is applied to the previous one with the intention of recovering phase. Thus, the end of the second gradient will compensate the application of the first. They will be, theoretically, phase rotating again. In practice, however, adjustment phase

is not perfect in all protons (Figure 3.5a), resulting in a loss of signal that can be picked up by the scanner. The main reason for this imperfect adjustment is the diffusion of water molecules (Brownian motion).

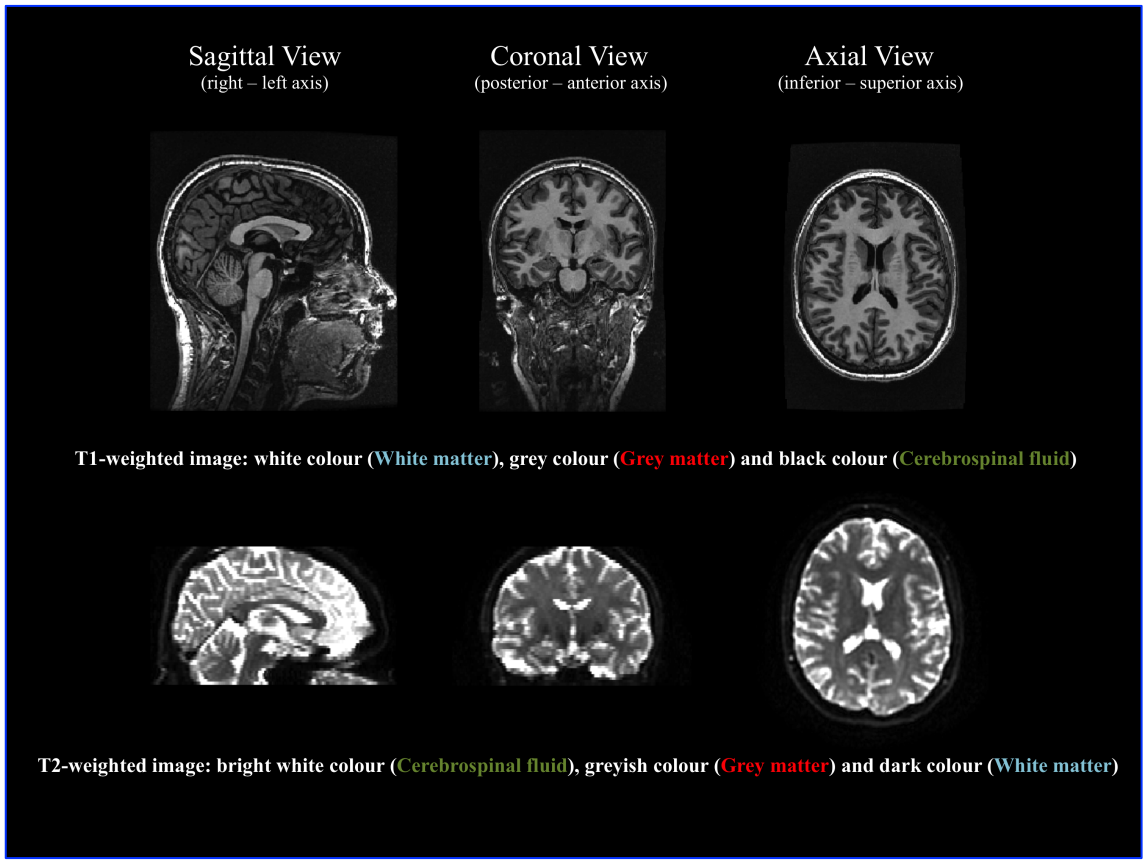


Figure 3.4. Top panel shows a T1-weighted image in three different views: sagittal, coronal and axial. Bottom panel shows an example of T2-weighted image in the same views.

Brownian motion is conditioned by the presence or absence of local barriers. The water movement is different when there are local barriers; for example, a molecule in a glass of water has an isotropic movement, i.e. with equal probability in any direction. However, in the presence of barriers, the molecules move preferentially in one direction of space; this movement is denominated anisotropic. For the brain, the hydrogen nuclei have an isotropic movement in the ventricles (cerebrospinal fluid), while the movement in the white matter structures is anisotropic, since the white matter structures are generally covered by myelin sheaths. Therefore, these structures act as physiological barriers. The movement of hydrogen in grey matter is also isotropic, but less so than for the cerebrospinal fluid.

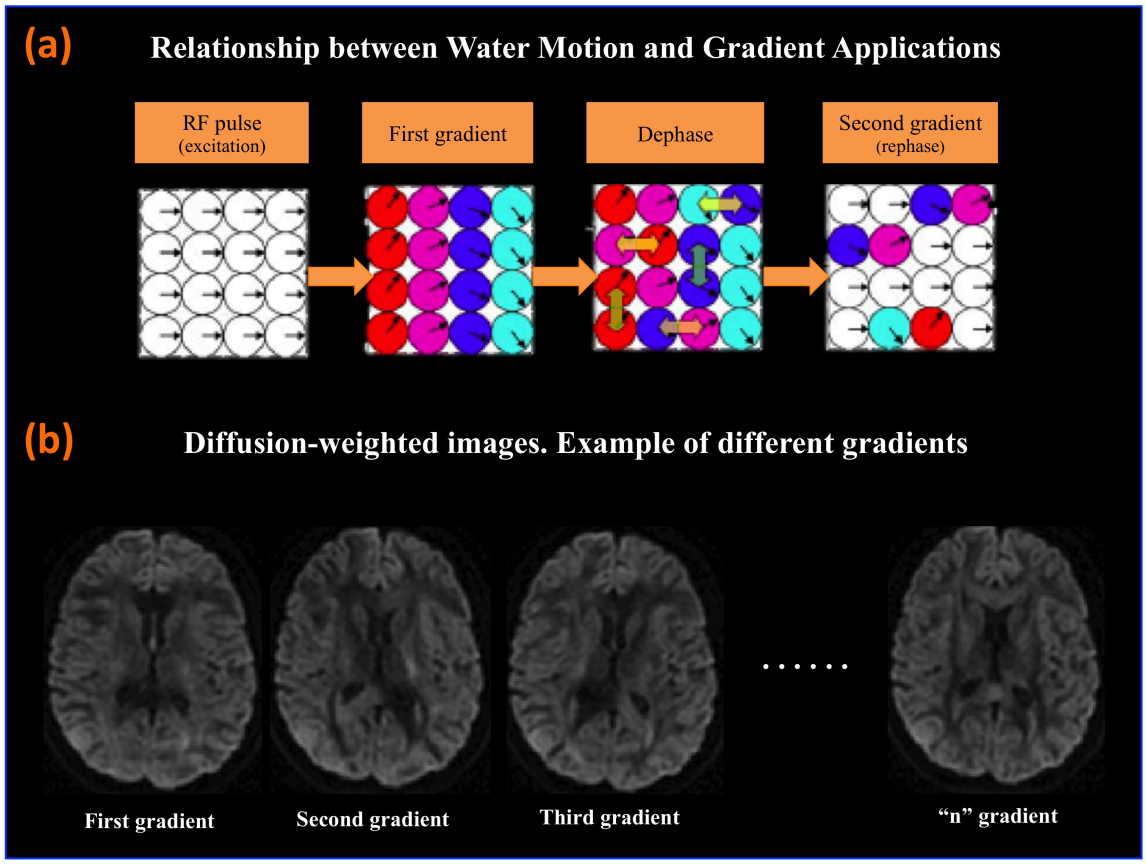


Figure 3.5. Top panel (a) shows a schematic Diagram to Explain the Relationship between Water Motion and Gradient Applications. The first gradient introduces differences in the precessional frequency (precess faster warmer colors), and the second cancels those differences; however, in the meantime some protons have moved (vertical and horizontal arrows in step dephase). Those that have moved in the same direction as the gradient (here, horizontal) derived in a signal loss that can be detected (diffusion). Note that, to detect signal loss "Vertical", we should apply a gradient in that direction. Adapted from Mori and Zhang (2006). Bottom panel (b) shows examples of diffusion-weighted image. Images are of the same coordinated and they are from the same individual, but each resulting from the application of gradients in different direction. Note the difference in contrast in different areas of the image, especially visible around the ventricles. Dark regions are the areas with greater loss of signal.

The application of dual gradient MRI allows capturing an image that is sensitive to the diffusion of water by detecting the phase mismatch (diffusion-weighted image). However, the gradient only can capture diffusion, which takes place in the direction in which it has gradient applied, which occurs along the longitudinal axis of the gradient. Therefore, information obtained on the three dimensional structure will be very limited. To improve the quality of the information obtained in a DWI, different gradients pulses are combined from different orientations (Figure 3.5b). The resulting image allows identifying the spread in all directions (or as many as gradients have been applied), i.e.

this image allows to detect and reproduce the white matter tracts and the computation of fractional anisotropic (FA) a scalar measure of the eccentricity of the ellipsoid of each point of the image (voxel) (Figure 3.6). An ellipsoid is a curved surface whose three main orthogonal sections are elliptical closed. FA is depicted in white (high anisotropy) / black (low anisotropy), or with a color code indicating the orientation: red (right/left), blue (top/bottom) and green (anterior/posterior).

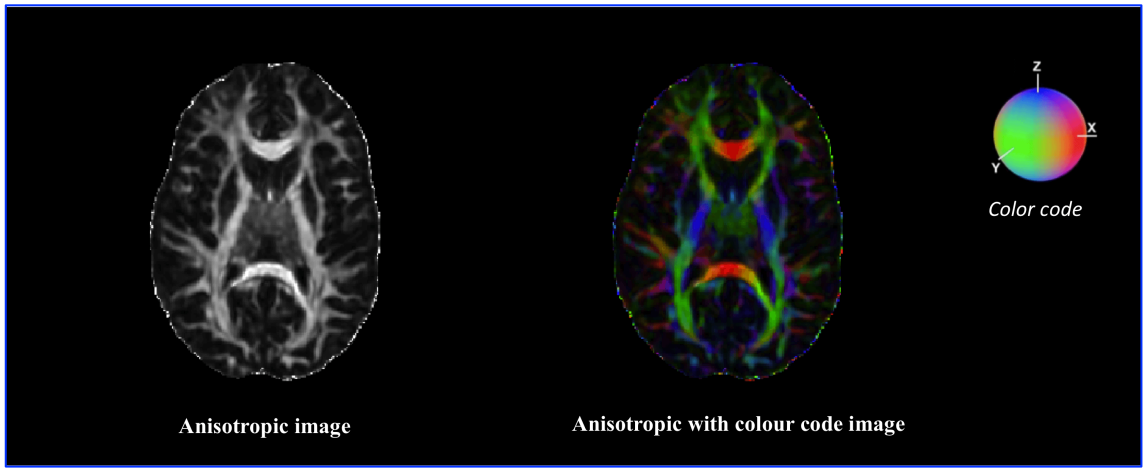


Figure 3.6. Left panel shows the anisotropy image of one subject, while right panel depicts the anisotropy map with color code. White regions depict zones with high anisotropy, while dark zones are the regions with low anisotropy.

The array elements of an MR image are known as "volume elements" or voxels in three dimensions (x, y, z). The information received by the RF reception coils is saved in each voxel, so that in each voxel the information received for a given spatial localization is represented. This information has different meanings depending on the kind of image (T1-weighted, T2-weighted, DWI, and so on).

3. 3. Surface-Based Morphometry.

Surface-Based Morphometry (SBM) approaches allow to process T1 weighted images for quantifying individual differences in brain anatomy. Cortical thickness, cortical surface area, grey matter volume, cortical curvature, or cortical complexity, are common outputs from the different SBM algorithms currently available.

Probably, the most studied macro-anatomical property of the brain is cortical grey matter volume (CGMV). Voxel-Based Morphometry (VBM) has been very popular for studying CGMV (Ashburner & Friston, 2000; Mechelli, Price, Friston, & Ashburner,

2005). However, VBM has some weaknesses, including its high sensitivity to relatively small differences in preprocessing (Thomas et al., 2009). Also, most importantly, VBM cannot disambiguate the two CGMV components: cortical thickness (CT) and cortical surface area (CSA). SBM creates surfaces representing structural boundaries (white matter–grey matter; grey matter–cerebrospinal fluid) using different algorithms (Fischl & Dale, 2000; Kim et al., 2005; Thompson et al., 2004), allowing the computation of several local grey matter measures, such as cortical thickness and cortical surface area. CT represents the number of cells within a vertical column, whereas CSA reveals the overall degree of folding (Rakic, 1988). GMV is a function of CT and CSA (Panizzon et al., 2009).

Recent research suggests that (1) CT, CSA and GMV are heritable indices; (2) CT and CSA are somewhat independent; (3) the variability in CSA is higher across individuals than in CT; and (4) the genetic substrate of CT and CSA is different (Colom et al., 2013b; Chen et al., 2013; Panizzon et al., 2009; Sanabria-Díaz et al., 2010; Vuoksimaa et al., 2014; Winkler et al., 2010). Analyses for CGMV will not be reported here since the correlation between CSA and CGMV is close to .90 (Colom et al., 2013b; Román et al., 2014; see Winkler et al., 2010).

There are different programs designed for SBM analyses (e. g., *FreeSurfer, CIVET, BrainSuite*, and so on) that vary in the algorithms used for processing the MR images. However, they share several steps required to process the data:

(1) Volumetric Spatial Normalization. The central goal here is to transform native brain scans to achieve spatial voxel-by voxel correspondence across subjects for group analyses.

(2) Correction of Intensity Non-Uniformity Artifacts (Intensity Normalization). Ideally, MR image contrast is due to brain tissue properties. However, there are different artifacts that may affect the MRI images. There are different methods to attenuate this problem.

(3) Tissue Segmentation and Classification. After the removing of non-biological signals, the images are divided according to biological tissues: grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF).

(4) Cortical Surface Reconstruction and Parametrization. After the classification of the different tissues types (grey matter, white matter and cerebrospinal fluid), 3D models per subject (typically triangulated meshes; see Figure 3.7) are computed for the GM-WM boundary (inner surface) and the GM-CSF boundary (outer or pial surface). Each triangle of the mesh is called a vertex, and it's the unity of analysis for SBM.

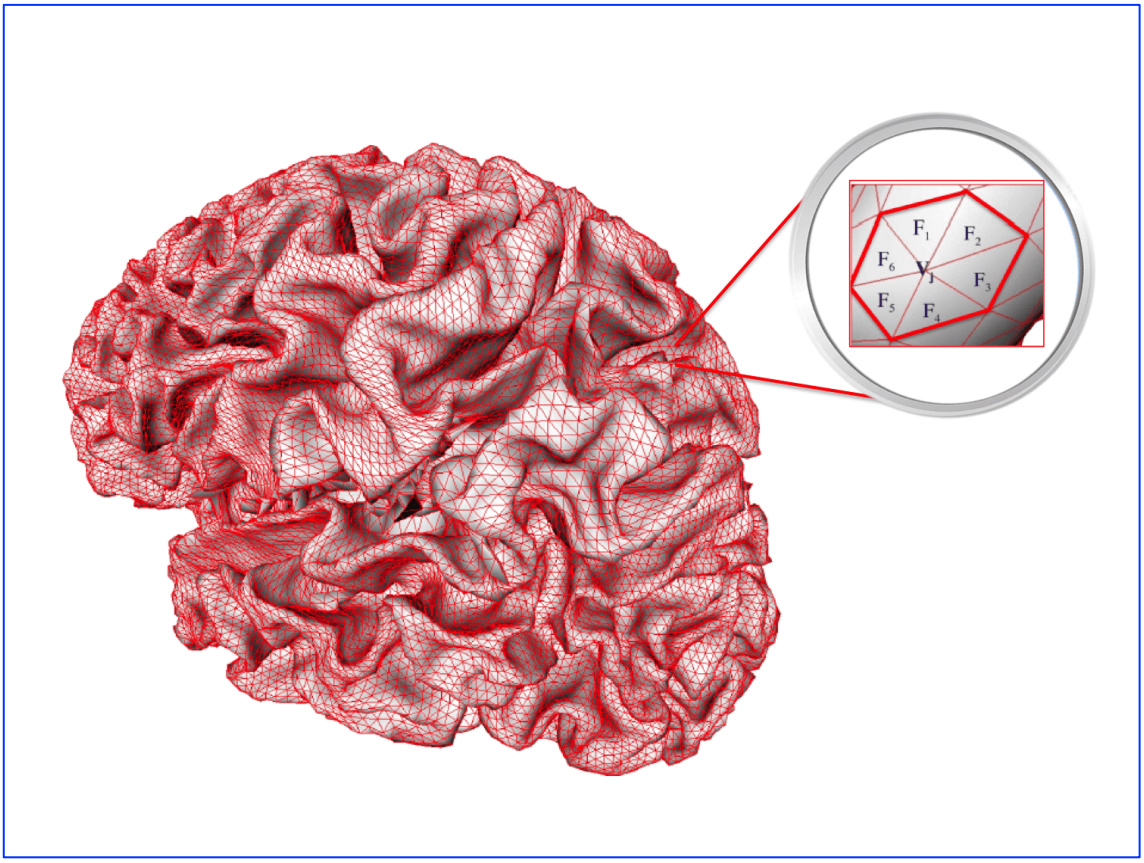


Figure 3.7. Example of triangulated meshes used in CIVET pipeline.

(5) Cortical Surface Registration. In this step, the individual meshes are matched while accounting for individual differences in folding and sulcal pattern.

(6) Computation of CT, CSA and GMV. These transformations are aimed at achieving anatomical correspondence across subjects at each spatial location of the mesh (vertices). Then, morphological indexes can be computed, but also in native space (e.g., by rescaling the aligned cortical surfaces back to native space dimension, i.e. original space of the subject). There are different methods for computing CT but mainly, cortical thickness is computed as the distance for each vertex between inner surface (GM-WM boundary) and pial surface (CSF-GM boundary). Cortical surface area is computed as the sum of the area of all faces that meet at a given vertex and divided by

three producing a measure of area per vertex. Finally, GMV is a combination of CT and CSA.

(7) Smoothing Morphometric Data. This is a mandatory step before running stats. It consists of a blurring of the computed data by convolving it with a three dimensional Gaussian kernel. After this process, the morphological indices at each vertex become a weighted average of the values corresponding to the surrounding vertices. There are several reasons for this step. Smoothing (a) has the effect of rendering the data more normally distributed, thereby increasing the validity of parametric statistical tests; (b) helps to compensate for the inexact nature of the spatial normalization; (c) allows the reduction of noisy variability due to the limited MRI sampling; and (d) reduces the number of comparisons necessary to correct for random field theory (Lerch & Evans, 2005), since the interdependence of the neighboring vertices is increased. These improvements are obtained at the cost of spatial resolution and hence a loss of local detail (Zhao, Boucher, Rosa-Neto, & Evans, 2013).

As noted above, even when all these steps are shared by most of the SBM protocols, there are important differences in the algorithms behind each step (Martínez, 2014; Martínez et al., 2015). In this work, we will use the CIVET pipeline (version 2.0) developed at the MNI. CIVET is a fully automated pipeline to extract and co-register cortical surfaces for each subject (Ad-Dab'bagh et al., 2006). The CIVET pipeline will be described later on Chapter 7.

3. 4. Tractography Methods for Reconstructing White Matter Structural Connectivity Matrices

This section is devoted to describing the processing workflows used to obtain structural connectivity matrices from diffusion-weighted and T1 weighted images. In standard analyses, these matrices contain information about the white matter pathways connecting different predefined grey matter regions. As mentioned above, the myelin sheaths hinder the isotropic movement of hydrogen nuclei within white matter fibers. Therefore, the hydrogen nuclei trace an anisotropic movement inside the white matter. The reconstruction of the movement of the hydrogen nuclei in each voxel of the MRI image is the first step to compute a structural connectivity matrix. The number of diffusion-weighted images depends on the number of gradient pulses employed during

the acquisition. A higher number of gradients improve the quality of the image, because a higher number of directions during the acquisition result in a better representation of the directionality of movement of hydrogen nuclei within voxels. Depending on the quality of this representation in the image, different methods may be employed for reconstructing both the directionality of the movement and the white matter trajectories.

Standard processing pipelines for building white matter structural connectivity matrices include two broad workflows that might run in parallel. One of them is aimed at reconstructing the white matter pathways from the diffusion weighted images, while the other works over the T1 or T2 weighted images for dividing the brain in nodes or regions of interest (ROIS). The final step is the computation of connectivity matrix.

(1) Processing diffusion-weighted images for estimating intra-voxel fiber orientations and for reconstructing the axonal nervous fiber trajectories. The two main methods for reconstructing the directionality of the movement are the diffusion-tensor imaging model (DTI; Basser, Mattiello, & LeBihan, 1994) and the high-angular-resolution diffusion imaging models (HARDI; Tuch et al., 2002a). As described below, HARDI models have several advantages over DTI, however a minimum number of gradients are necessary to employ HARDI models. Tournier, Calamante, and Connelly (2013) found that 45 directions are sufficient to fully characterize the DWI signal.

The DTI model estimates a symmetric tensor representing water molecular motion along the three coordinate axes (x, y, z). The tensor of diffusion is computed adjusting to a tridimensional ellipsoid the different diffusion measures acquired. This tensor is geometrically defined by an ellipsoid (see Figure 3.8- Left). The main axis of the ellipsoid is an estimation of local major fiber orientations on white matter tissues (see Figures 3.8- Right and 3.9-Left). Due to the limited number of reconstructed directions (axes), this model fails to estimate the trajectory of the water molecules in voxels where there are crossing or kissing fibers, that is, fibers with different orientations.

In order to overcome this limitation of the DTI models, HARDI models represent the distribution of water diffusion along a larger number of directions, leading to a more detailed mapping of multiple, complex intravoxel fiber orientations. Different reconstruction approaches have been proposed, although qball imaging (QBI; Canales-Rodríguez, Melie-García & Iturria-Medina, 2009; Tuch, 2002b) and diffusion spectrum

imaging (DSI; Wedeen et al., 2008) have generated considerable attention due to a number of benefits such as model independence, theoretical soundness, and the ability to resolve intravoxel orientation heterogeneity (Canales-Rodríguez et al., 2009). Both approaches are based on the computation of the orientation distribution function (ODF), which is obtained by means of the radial projection of the diffusion propagator in DSI models, while in QBI models is directly estimated from the data without computing the diffusion propagator. In general, ODF characterizes the distribution of water diffusion along different directions on a sphere. Therefore, a better reconstruction of the tracts is obtained since this function allows for the estimation of more than one main direction at each voxel in the DWI image (see Figure 3.9 - Right).

Once the directionality of the water molecule movement is represented, the data is ready for tractography algorithms. Deterministic and probabilistic approaches are used as methods of reconstructing fiber trajectories. Deterministic tractography is based on the use of intravoxel seeds and the reconstruction of the tract according to which fiber orientation can be estimated. Therefore, deterministic algorithms can be used to trace long axonal pathways (Conturo et al., 1999; Jones, Simmons, Williams, & Horsfield, 1999; Mori, Crain, Chacko, & Van Zijl, 1999; Basser, Pajevic, Pierpaoli, Duda, & Aldroubi, 2000). On the other hand, the probabilistic tractography (Behrens et al., 2003, 2007; Parker & Alexander, 2003; Parker, Haroon, & Wheeler-Kingshott, 2003) makes inferences of nervous fiber orientations from the potential paths from the seed point using probabilistic algorithms. Recently, global algorithms have emerged to investigate the most probable fiber trajectories that maximize a global goodness-of-fit criterion (Iturria-Medina et al., 2007; Sherbondy, Dougherty, Ananthanarayanan, Modha, & Wandell, 2009; Tuch, 2002b).

(2) Node definition. The first step is the selection of an atlas, which will be employed to make a parcellation of grey matter of the MRI image. This issue is highly relevant since the selection of the atlas has a direct influence on the network (Zalesky et al., 2010a). After, grey matter tissue is divided according to independent regions delimited by the selected atlas.

(3) Finally, an **anatomical connectivity matrix** is created by combining the structural information contained on the estimated fiber trajectories and the considered grey matter parcellation. The matrix contains the same number of rows and columns (the number is

defined by regions of the atlas). Each value of the matrix depicts the connectivity between a node defined in row "x" with a node defined in column "y". The main diagonal shows the connection of the node with itself. The different connectivity values should be defined trying to capture the physiological properties of the underlying connections and/or the evidence supporting the existence of each connection (Iturria-Medina, 2013).

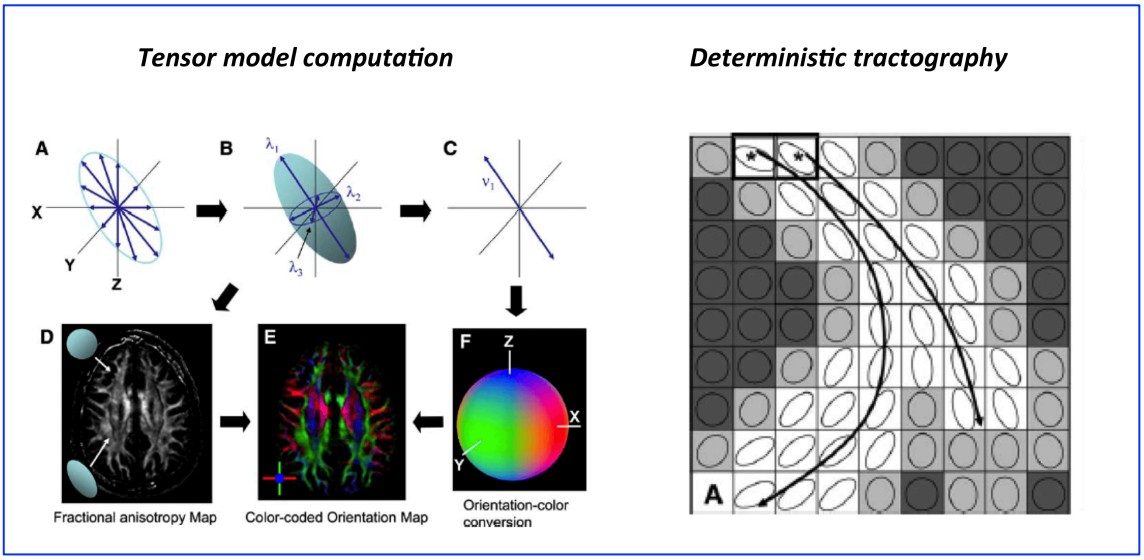


Figure 3.8. The principle of DTI are depicted in the right of the image: (A), the shape and the orientation of a ''diffusion ellipsoid'' is estimated (B). This ellipsoid represents the movement of the water molecules inside a determinate voxel, then (D) an anisotropy can be created from the shape. Dark regions are isotropic (spherical) and bright regions are anisotropic (elongated). From the estimated ellipsoid (B), the orientation of the longest axis can be found (C), which is assumed to represent the local fiber orientation. This orientation information is converted to a color (F) at each pixel. By combining the intensity of the anisotropy map (D) and color (F), a color-coded orientation map is created (E). The left of the figure depicts a schematic diagram of a basic algorithm for tract reconstruction (A). Average fiber orientation is estimated from diffusion anisotropy at each pixel, and a line is propagated from a pixel of interest (pixels with asterisks) following the fiber orientation, until it reaches a brain region of low anisotropy (dark pixels). Adapted from Mori and Zhang (2006).

In this work, we will employ a parcellation based on the *FreeSurfer* atlas, but with some modifications for the temporal lobe. Specifically, the OASIS template (Klein & Tourville, 2012) will be used here. As a method of reconstruction of intravoxel orientation, we will employ the HARDI approach based on q-ball space. Finally, as a method of reconstruction of the fibers, we will employ a global probabilistic algorithm based on graph tractography (Iturria-Medina et al., 2007). Specific details are described on Chapter 8.

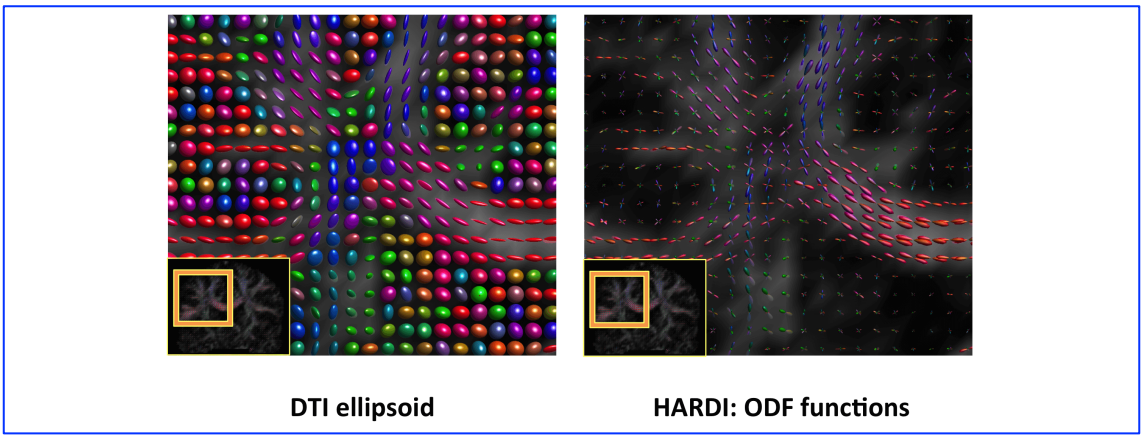


Figure 3.9. DTI ellipsoids (a) and HARDI model (b) for a coronal section showing lateral projections of the corpus callosum (left-right: red ODF lobes) crossing through the fibers of the corona radiata (inferior-superior: blue ODF lobes) and the fibers of the superior longitudinal fasciculus (anterior-posterior: green ODF lobes). DTI is unable to resolve these crossings. Adapted from Jeurissen (2012).

There are different ways to analyze connectivity matrixes: Link-based, ROI-based, graph theory, Network-based Statistic (NBS), etc. Methodology employed in this report is based on graph theory indices and NBS.

Graph-theory allows the computation of different indices for each connectivity matrix. These are some examples (Dennis & Thompson, 2013; Iturria-Medina, Sotero, Canales-Rodríguez, Alemán-Gómez, & Melie-García, 2008).

(1) ***Characteristic path length*** is an average measure of the path length in a network. This index explains how easy it is to move from one node in a network to any of the other nodes, regardless of the physical length of the axons, considering only what is connected. ***Lambda*** is a value considered as a normalized version of the characteristic path length index using randomly created networks of the same size and complexity.

(2) ***Mean clustering coefficient*** is a measure of how many neighbors of a given node are connected to each other. The normalized version of this index is named ***Gamma***.

(3) ***Ratio*** (***Gamma/Lambda***) is used to represent "small-worldness" properties, since it symbolizes the balance between network differentiation and integration. Ratios higher than 1 mean that the network support small-worldness (networks highly connected and with short average steps for connecting different sub-networks).

(4) ***Global efficiency*** is considered the inverse of characteristic path length since networks with a lower characteristic path length are more efficient. ***Regional efficiency*** is the efficiency calculated on nodal neighborhoods. ***Local efficiency*** indicates how well the information is communicated within the neighbors of a given node when this node is removed.

(5) ***Degree*** reflects the number of nodes to which a given node is connected in a binary network, while ***Strength*** is the degree in a weighted network.

(6) ***Modularity*** is the degree to which a system may be subdivided into smaller networks.

Figure 3.10 depicts the meaning of these indices.

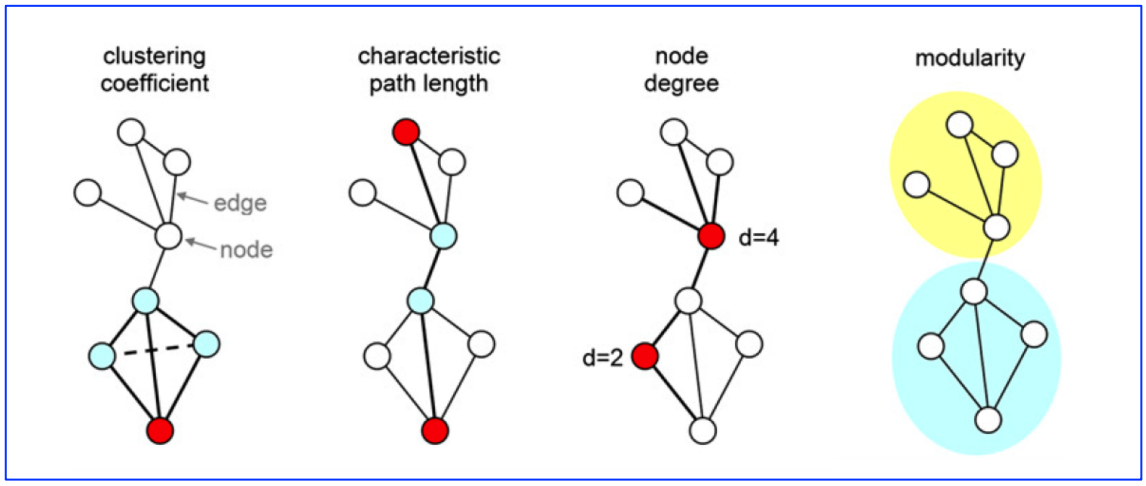


Figure 3.10. Graphic example of measures that can be computed from a structural connectivity matrix, adapted from Tymofiyeva et al., (2014).

On the other hand, NBS is intended to control family-wise error rate (FWE) when univariate testing is computed at every connection comprising the graph (Zalesky, Fornito, & Bullmore, 2010b). NBS specifies information about whole-brain structural organization separating sets of highly interconnected regions (sub-networks) instead of paired regions. This approach has been applied for comparing healthy individuals and patients (Bai et al., 2012; Verstraete et al., 2011; Zhang et al., 2011). Only recently it has been applied for comparing groups of healthy participants (early bilinguals and monolinguals) (García-Pentón, Fernández, Iturria-Medina, Gillon-Dowens & Carreiras, 2014).

3. 5. Brain Development: Spontaneous changes across the life span.

MRI techniques allow the in-vivo study of age-related changes in biological indices. Available evidence suggests that (a) the brain is a dynamic system; (b) networks in the brain are structurally flexible; and (c) these networks are modified across the lifespan (Caroni, Donato, & Muller, 2012). Therefore, there are remarkable spontaneous changes in brain structure due to normal development (Dennis & Thompson, 2013). The consideration of these developmental changes is relevant for the current work, since brain changes in a passive-control group will be compared with those of a trained group.

3.5.1. *Grey matter development.*

Grey matter consists of neuronal cell bodies, dendrites, short axons, glial cells, and blood vessels. Maturational changes in grey matter have been widely studied (e.g. Casey, Tottenham, Liston, & Durston, 2005; Giedd et al., 1999; Giedd & Rapoport, 2010; Raznahan et al., 2011; Sowell et al., 2003). However, these changes might vary for distinguishable indices, such as cortical grey matter volume, cortical thickness, cortical surface area, or gyrification (Raznahan et al., 2011).

Grey matter volume is the most studied measure. Changes on this index are non-linear and show decrements across the life span (Gogtay & Thompson, 2010; Sowell et al., 2003; Wierenga, Langen, Oranje, & Durston, 2014), although it can show an inverted U-shaped curve in frontal, parietal and temporal areas (Casey et al., 2005; Giedd et al., 1999; Giedd & Rapoport, 2010). Motor and sensory systems mature earlier, followed by temporal and parietal association regions. Higher-order association areas (prefrontal and lateral temporal cortices) are the latest to reach maturation (Cassey et al., 2005; Giedd et al., 2010; Gogtay & Thompson, 2010).

Several hypotheses have been proposed to explain the differential trajectories of grey matter volume across life: (a) synaptic and neuronal pruning (Paus, 2005); (b) neuronal atrophy and reduction of dendritic arborization (Hof & Morrison, 2004); and (c) increased proliferation of myelin into the periphery of the cortical neuropil, which would change the MR signal from grey matter in the younger subjects to white matter in the older subjects (Raznahan et al., 2011). However, as detailed above, grey matter

volume combines two relatively independent cortical measures: cortical thickness (CT) and cortical surface area (CSA).

Raznahan et al., (2011) studied the trajectories of different grey matter indices (cortical grey matter volume, cortical surface area, cortical thickness, and gyrification) in a longitudinal study. Their main findings are shown in Figure 3.11.

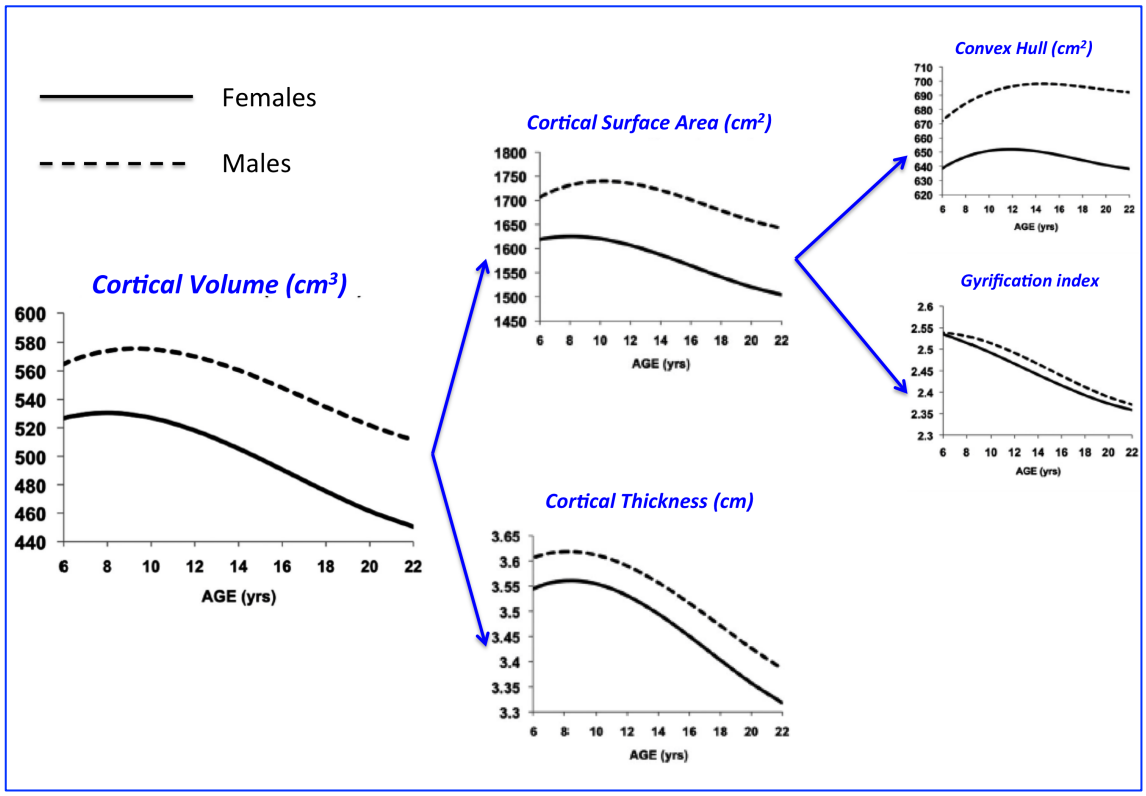


Figure 3.11. Developmental trajectories for cortical volume and each of its subcomponents in males and females (adapted from Raznahan et al., 2011).

The developmental curve for total cortical grey matter volume was U inverted. This pattern was found for CT, and CSA, whereas the trajectory for GI was a lineal decrease. Similar results were found for cortical thickness in other reports (e.g. Shaw et al., 2006). However, Shaw et al. (2008) found that these trajectories for CT are not homogenous across the brain, since linear, quadratic or cubic trajectories were observed in different areas. They argued that the complexity of the laminar architecture could explain these different trajectories. However, other studies showed CT have monotonic decrease trajectories (e.g. Burgaleta, Johnson, Waber, Colom, & Karama, 2014; Zhou, Lebel, Treit, Evans, & Beaulieu, 2015). For cortical surface area, results are also mixed across studies: decrements with age after childhood (e.g. Raznahan et al., 2011; Schnack et al.,

2014), increments (Zhou et al., 2015), and increments in some regions and decrements in other regions (Burgaleta et al., 2014).

In sum, brain volume increases throughout the first years of life and then remains relatively stable. This stability might be attributed to white matter volume increments until the fourth/fifth decade, which compensates for grey matter decrements (Raznahan et al., 2011; Rivkin, 2000).

3.5.2. *Development of white matter.*

White matter comprises myelinated axons. In contrast to grey matter, white matter only stops its volume increments at the fourth or fifth decade of life (Cassey et al., 2005; Giedd et al., 1999; Giedd & Rapoport, 2010; Sowell et al., 2003). However, Tamnes et al. (2010) found nonlinear increments in at least one-third of the white matter tracts analyzed. Increments in white matter are explained by a progressive myelination of the axons by oligodendrocytes enhancing neuronal conduction and communication (Cassey et al., 2005; Paus, 2010) up to middle age (40/50 years old). Decrements are explained by an accumulation of microlesions and demyelination (Davis et al., 2009).

The study of white matter trajectories is highly relevant, since white matter tracts connect different grey matter regions. As has been explained before, connectivity matrices are employed to show the connection between white matter and several grey matter regions. The main outcomes for the studies that used the graph theory approach to study brain development are summarized in Table 3.1.

The general pattern denotes an increment in the first years followed by a decrement after adolescence-early adulthood. Consequently, dynamic changes in connectome organization throughout the life span are suggested to follow an inverted U-shaped pattern (Collin & van den Heuvel, 2013). This decrement of the indices is probably caused by weakening of some short-range connections and strengthening of long-range connections (Casey et al., 2000) as a consequence of synaptic pruning during adolescence (Gogtay et al., 2004; Paus, 2005), and increases in intracortical myelination in young adulthood (Giedd et al., 1999).

Table 3.1. Summary of studies that researched the development of brain structural connectivity. The studies are organized according to the increasing age of the sample population; yrs = years

Study	*N*	*Ages*	*Indices*	*Outcomes*
Yap et al., (2011)	39	2 weeks, 1 yrs, 2 yrs	Global efficiency Local efficiency	Constant Increment
Fan et al. (2011)	28	1 month - 2 yrs	Small-worldness Modularity	Increment Increment
Hagmann et al. (2010)	30	18 months - 18 yrs	Strength Global efficiency Clustering coefficient	Increment Increment Decrement
Dennis et al. (2013)	439	12 yrs- 30 yrs	Path length Clustering coefficient Gamma Small-worldness Modularity	Decrement Decrement Decrement Decrement Decrement
Gong et al. (2009)	95	19 yrs -85 yrs	Overall connectivity Local efficiency	Decrement Decrement

In summary, most cortical grey matter indices show a non-lineal inverted U-shape trajectory (see Cassey et al., 2005). The maturation in each region is disparate, since motor and sensory areas have their highest point at earlier ages than regions associated to higher cognitive domains (Gogtay & Thompson, 2010). Also, the reduction is more pronounced during adolescence and it is smoothed during adulthood (Sowell et al., 2003; Zhou et al., 2015). This pattern of results has been found for CT, grey matter volume indices, and structural connections between grey matter regions through white matter tracts. However, results for CSA are mixed (Wierenga et al., 2014; Zhou et al., 2015). On the other hand, white matter volume increases across the life span balancing out the decrement in grey matter (Figure 3.12). Relevant changes are governed by the genetic makeup (Barbey et al., 2014b; Brans et al., 2010), and, also, developmental changes in CT and CSA are modulated by baseline differences in intelligence (Burgaleta et al., 2014; Ramsden et al., 2011; Schnack et al., 2014; Shaw et al., 2006).

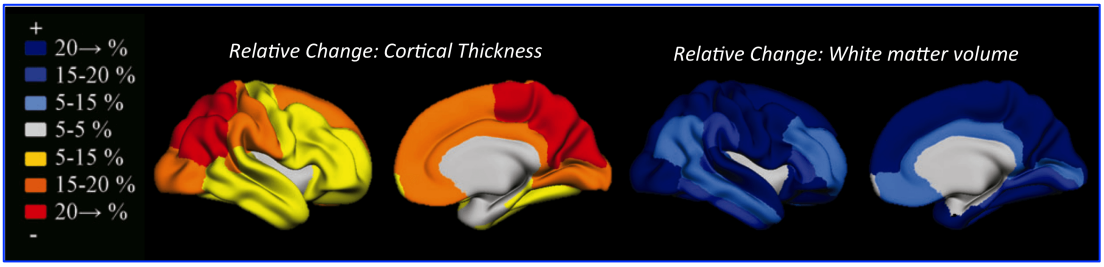


Figure 3.12. Relative magnitude of developmental changes over the age span 8--30 years. Percentage change of cortical thickness and white matter volume. Hot colors represent decrement with ages, while cold colors depict increment across life span (adapted from Tamnes et al., 2010).

3. 6. Network shared by intelligence and working memory.

Before summarizing the studies that focused on brain changes after cognitive training, we would like to point out that a working memory task (n-back task) is employed as a training task in the present work. This is relevant because one of the main ideas is that working memory and fluid intelligence share cognitive components and biological networks. Therefore, understanding the networks shared by fluid intelligence and working memory is crucial to determining the regions where some if not most a priori biological changes are expected.

3.6.1. *Intelligence.*

3.6.1.1. P-FIT

The Parieto-Frontal Integration Theory (P-FIT) of intelligence (Jung & Haier, 2007) is based on a review of structural and functional studies relevant to intelligence (Figure 3.13). Highlighted brain regions are organized in stages according to their presumed main psychological function.

First stage: Processing of sensory information. Occipital and temporal areas support this stage. The extrastriate cortex (Brodmann areas –BAs – 18 and 19) and the fusiform gyrus (BA 37) are involved with recognition, imagery and elaboration of visual inputs, as well as Wernicke's area (BA 22), which is focused on analysis and elaboration of syntax of auditory information.

Second stage: Symbolism, abstraction and elaboration. All regions of this stage are placed in the parietal lobe: BAs 39 (angular gyrus), 40 (supramarginal gyrus), and 7

(superior parietal lobule). Therefore, the information is processed in the posterior regions of the brain (stage one), and then, in the parietal areas, the elaboration of the information is carried out.

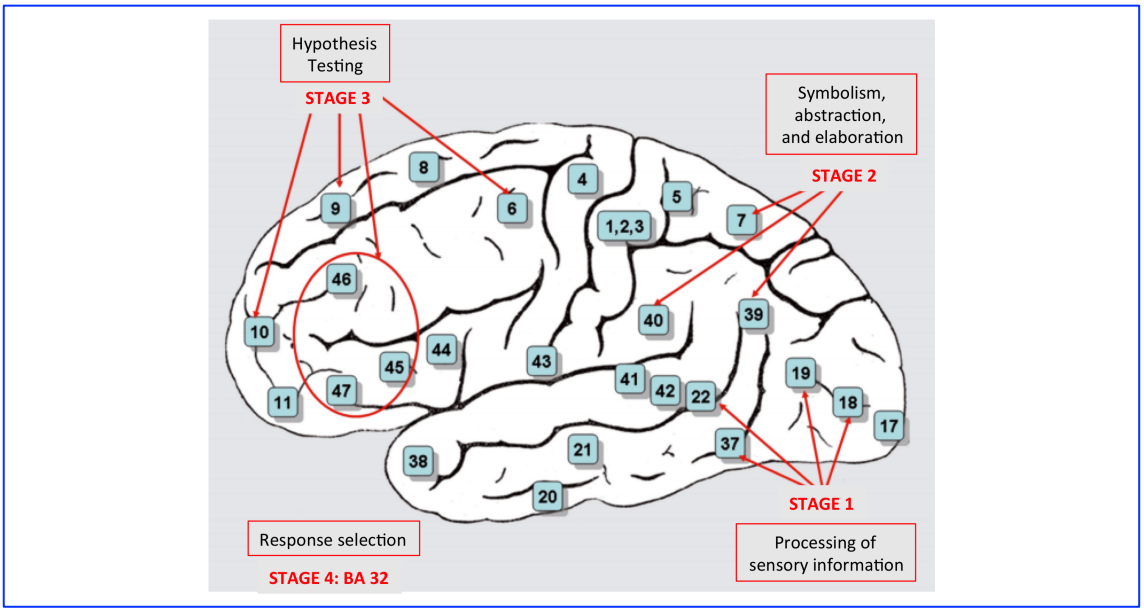


Figure 3.13. Processing stages proposed by the P-FIT model, adapted from Colom, Karama, Jung, & Haier (2010).

Third stage: Hypothesis testing. In this phase, the parietal areas interact with the frontal regions (BAs 6, 9, 10, 45, 46, and 47) enabling problem solving, evaluation, and hypothesis testing.

Finally, the ***fourth stage is devoted to response selection*** and inhibition of alternative responses, once the best solution is determined in the previous stage. This stage is carried out in the anterior cingulate (BA 32).

Jung and Haier (2007) proposed that regions of the dorsolateral prefrontal cortex (BAs 9, 45, 46, and 47) and the parietal cortex (BAs 7 and 40) should be considered the most important for human intelligence. However, Colom (2007) noted that there is a high variability among studies: only a very small number of discrete brain areas approach 50% of convergence across published studies employing the same neuroimaging strategy. In the case of structural studies, 32 brain areas were nominated, but only BAs 39-40 and 10 approach 50% of convergence. These discrepancies might be attributed to (a) the sample tested (e.g. sex, age, and lesion vs. healthy subjects), (b) the intelligence construct considered (e.g. the *g* factor, fluid, crystallized, or visuo-spatial intelligence;

see Román et al., 2014) and the psychometric tests used to measure these constructs (e.g. there are more clusters and voxels related to grey matter volume with increased *g* loadings of the administered measures; see Colom, Jung, & Haier, 2006), (c) the specific neuroimaging processing protocol employed (see Martínez et al., 2015), and (d) the measured brain property (for a review see Colom & Thompson, 2011).

Basten, Hilger and Fiebach (2015) conducted a voxel-based quantitative meta-analysis of 12 structural and 16 functional human brain imaging studies to exam the P-FIT model. The outcomes for structural studies showed 12 clusters distributed in lateral and medial frontal, temporal, and occipital cortices, as well as in subcortical structures. The functional meta-analysis resulted in eight clusters distributed across both hemispheres, located in lateral frontal, medial frontal, parietal, and temporal cortices. No overlaps were found between structural and functional results (see Figure 3.14).

Results of this meta-analysis for structural and functional studies showed a reasonable match with the P-FIT. However, parietal lobe only was found as statistically significant in functional studies, while occipital and subcortical only were found in structural studies.

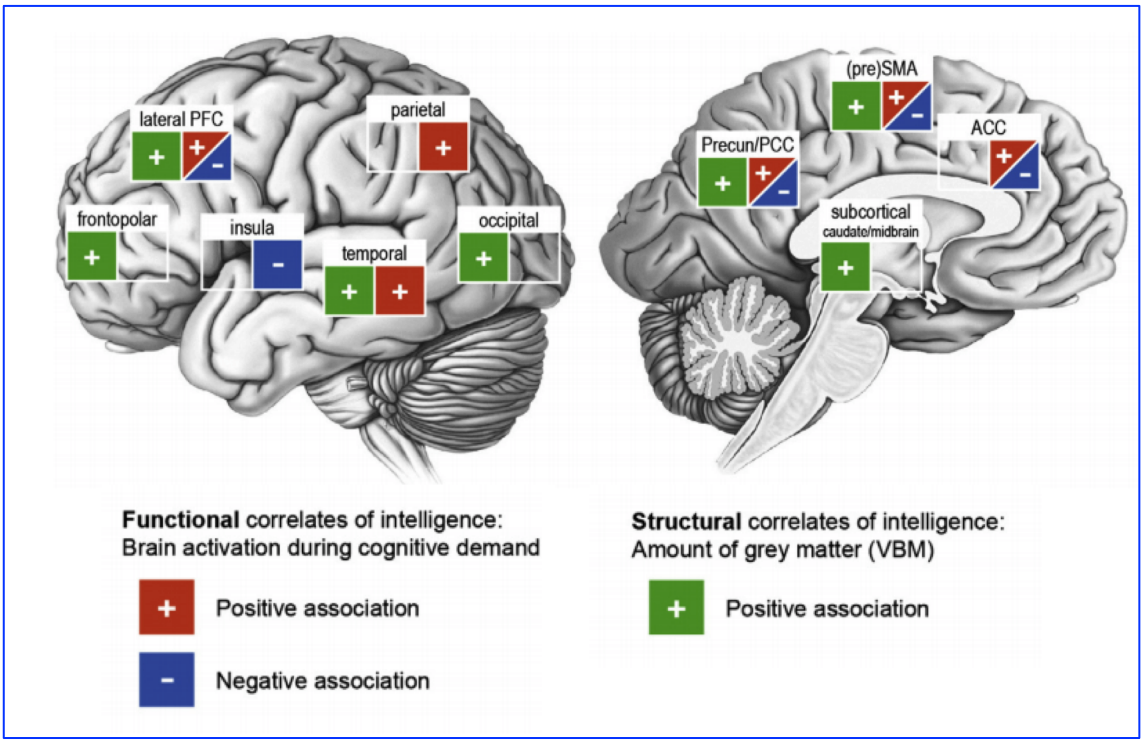


Figure 3.14. Outcomes of the meta-analysis of Basten al., (2015) with structural and functional data.

3.6.1.2. Cortical thickness

Shaw et al., (2006) analyzed the correlations between CT and IQ in a sample of 307 healthy children. They detected modest and non-significant positive correlations throughout most of the frontal, parietal and occipital cortex, and similarly non-significant negative correlations in the anterior temporal cortex. However, the value of this correlation changed from early childhood to late childhood.

In 2009, Karama et al. found positive correlations between CT and a general cognitive ability factor for a sample of 6 to 18 year olds. Positive correlations were found across the brain in young children (6-11 years old) and adolescents (12-18 years old). In 2011, Karama et al. extended this work calculating a *g* factor through a comprehensive neuropsychological test battery with an equivalent sample. Again, the correlations were positive and distributed across the brain. Furthermore, findings for specific factors vanished once the *g* factor was removed.

Narr et al., (2007) found positive correlations between IQ and cortical thickness in prefrontal (anterior-ventral prefrontal and frontopolar cortices; BA 10/11 and 47) and temporal cortices (inferior temporal, fusiform, and parahippocampal cortices; BA 20, 37, and 36) bilaterally. Choi et al., (2008) or Joshi et al., (2011) found similar patterns of correlations, but disparate results were observed in other studies (Colom et al. 2013b; Fleischman et al. 2013; Tamnes et al. 2011; Vuoksimaa et al., 2014).

3.6.1.3. Cortical surface area.

The number of studies addressing the relationship between intelligence and cortical surface area is much smaller (Colom et al., 2013b; Fleischman et al. 2013; Román et al., 2014; Vuoksimaa et al., 2014; Yang et al., 2013). These studies have found higher correlations for CSA than for CT. Román et al. (2014) argued that CSA might be a better index than CT for finding correlations between grey matter indices and intelligence, since: (a) CSA captures the balance between local specialization and global integration in the brain (Sanabria-Diaz et al., 2010), (b) CSA usually displays more variability than CT (Winkler et al., 2010), which may impact statistical sensitivity, and (c) CSA appears to be slightly more genetically determined than CT (Panizzon et al., 2009), a pattern that fits well with the fact that neurogenesis and neuronal migration—

two processes that affect CSA—are complete by the term of the gestation process (Hill et al., 2010). Given that intelligence is known to be heritable to a substantial degree (see, for instance, Nisbett et al., 2012), the association between intelligence and CSA might reflect those neural substrates of intelligence that are less sensitive to experience. Vuoksimaa et al., (2014) argued that an emphasis on neocortical surface area, rather than thickness, might be more fruitful for elucidating neocortical intelligence associations and identifying specific genes underlying those associations.

3.6.1.3. White matter

The relationship between human intelligence and the integrity of white matter has been much less investigated. Schmithorst, Wilke, Dardzinski, & Holland (2005) analyzed the relationship between intelligence and white matter structure using fractional anisotropy (FA) and mean diffusivity (MD) indices, finding positive correlations bilaterally for FA in white matter association areas (frontal and parieto-occipital areas). Yu et al., (2008) computed correlations between the integrity of several tracts (corpus callosum, cingulum, uncinate fasciculus, optic radiation, and corticospinal tract) and intelligence. The results showed that high intelligence participants do have greater white matter integrity than average intelligence participants in the right uncinate fasciculus. Pol et al. (2006) analyzed the heritability of focal white matter with a sample of 54 identical and 58 fraternal twin pairs, along with 34 of their siblings, finding that intelligence shares a common genetic origin with superior occipito-frontal, callosal, and left optical radiation. Finally, Chiang et al., (2009) have reported the very first study combining a genetically informative design and a DTI approach. They have studied 23 pairs of identical twins and 23 pairs of fraternal twins. They applied structural equation models at each point in the brain with the scores of white matter integrity. Afterwards, they generated a tree dimensional map of heritability. White matter integrity was found to be under powerful genetic control in bilateral frontal, bilateral parietal, and left occipital lobes. FA measures were correlated with the estimate of general intelligence and with non - verbal intelligence in the cingulum, optic radiations, superior fronto - occipital fasciculus, internal capsule, callosal isthmus, and the corona radiata. Further, common genetic factors mediated the correlation between intelligence and white matter integrity. This latter finding suggested a common physiological mechanism and a common genetic determination (Colom & Thompson, 2011).

3.6.2. *Working Memory.*

Most neuroimaging studies have addressed the functional correlates of working memory. In this regard, the meta-analysis by Wager and Smith (2003) concluded that: (a) there is an important dissociation of activation in the posterior cortex for task with verbal and spatial items, supporting the dorsal/ventral distinction (see Valdés-Conroy, Sebastián, Hinojosa, Román & Santaniello, 2014), (b) the left hemisphere is more activated than right hemisphere in verbal working memory tasks when the demands are low, but when executive demand increases, there is a right lateralization in the frontal cortex for spatial tasks and (c) working memory tasks that require executive processing produce more dorsal frontal activation than tasks based on storage only. Features of the tasks have a direct influence on the regions activated; for example, BAs 6, 8, and 9 (superior cortex) are involved in tasks where temporal order and updating processing are relevant. However, right BAs 10 and 47 (ventral frontal cortex) respond more frequently with demand for manipulation, and BA 32 (medial prefrontal cortex) is recruited when selective attention was required for information storage in working memory. Moreover, BA 7 (posterior parietal cortex) is involved in all types of executive function.

As discussed by Toga & Thompson (2005), unlike structural imaging, the interpretation of functional imaging data is closely dependent of the completed task. Within the framework of this thesis, the n-back task has been employed during the training program because it is one of the most popular measures of working memory in neuroimaging (Conway et al., 2005). The possibility of manipulating working memory load (e.g., 1-back, 2-back, 3-back, etc) and its response requirements account for the wide use of this task (Conway, Kane, & Engle, 2003).

Owen et al. (Owen, McMillan, Laird, & Bullmore, 2005) published a meta-analysis on the studies employing this task in functional neuroimaging contexts. Their conclusions were that discrete frontal (lateral premotor cortex, dorsal cingulate, medial premotor cortex; dorsolateral prefrontal, ventrolateral prefrontal cortex and frontal poles) and parietal (medial and lateral posterior) regions were robustly activated during n-back task performance.

The summary of neuroimaging studies (NeuroSynth) reported by Yarkoni, Poldrack, Nichols, Van Essen and Wager (2011) revealed regular activations in posterior parietal and anterior prefrontal regions when completing tasks requiring cognitive control, as is the case for the n-back task (Jaeggi et al., 2010b). Recently, Yeo et al., (2014) extended the previous meta-analysis applying a hierarchical Bayesian model algorithm to 10,499 experimental contrasts. This procedure allows considering principles of functional organization not easily captured by prior meta-analyses. Results for probabilistic relationships among tasks, cognitive components, and brain regions are shown in the next website:

http://www.ece.nus.edu.sg/stfpage/ybtt/papers/2014Brainmap/Interactive/index.html

Figure 3.15 shows the results of this meta-analysis for the n-back task.

The component associated with the n-back task are mainly located in bilateral dorsoprefrontal, bilateral superior frontal, right caudate anterior, bilateral inferior parietal, left middle temporal, and right precuneus.

Tang et al., (2010) reported negative correlations between intelligence measures and neuroimaging activation when performing the n-back task (the greater the intelligence the lower the activation), a result that supported the efficiency model of brain function (Haier, 1993; Haier et al., 1988; see Neubauer & Fink, 2009).

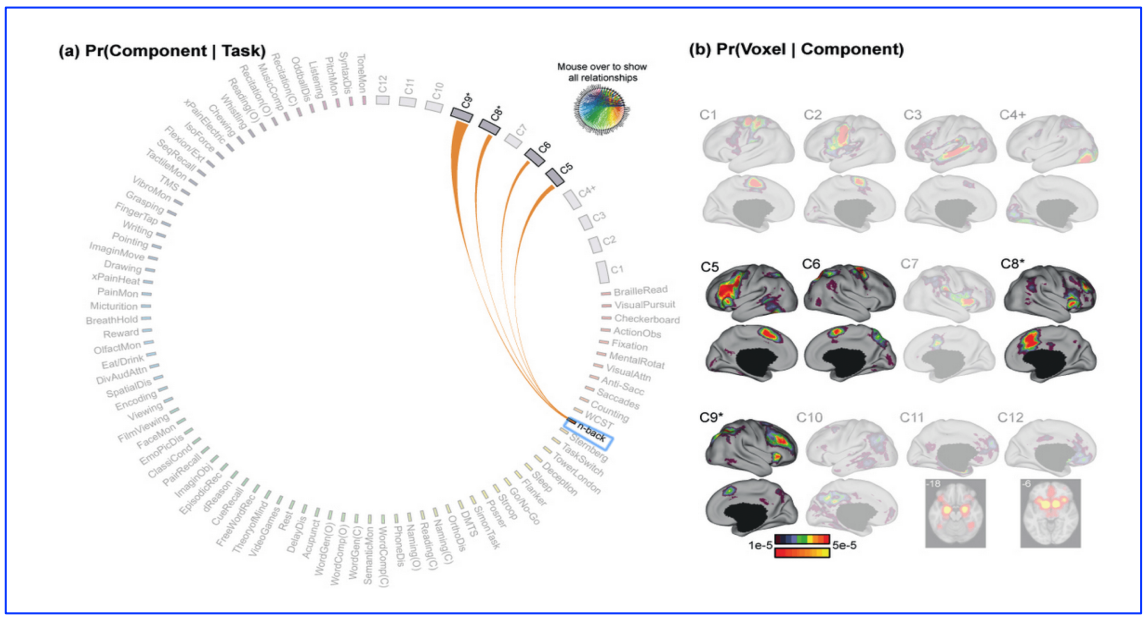


Figure 3.15. Results for the n-back task after Yeo et al.' meta-analysis (Yeo et al., 2014).

3.6.3. *Brain networks shared by intelligence and working memory.*

The neuroanatomical overlap between the general factor of intelligence and working memory capacity has been studied using different neuroimaging techniques. Colom, et al. (2007) applied a VBM approach, finding a common neuroanatomic frontoparietal network (Jung & Haier, 2007). These results were interpreted as suggesting that frontal regions support processes dedicated to cognitive control, whereas parietal areas are responsible for the capacity limitations of the human brain (Cowan, 2005).

Colom et al., (2013b) analyzed the neuroanatomic overlap between fluid intelligence (Gf), crystallized intelligence (Gc), visual-spatial intelligence (Gv), working memory capacity (WMC), attention control (ATT) and processing speed (Speed) measured at latent level. They applied VBM (see Figure 3.16a) and SBM (see Figure 3.16b) techniques, finding clusters shared by Gf, Gc and WMC in the frontal lobe.

Lesion studies allow to establish whether the frontal and parietal regions are necessary for these cognitive mechanisms (Barbey et al., 2014a). Barbey et al., (2014a). These type of studies have found that fluid intelligence and spatial working memory share clusters located in right frontal and parietal lobe (see Figure 3.16c), using structural equation models to defining the cognitive factors (see also Gläscher et al., 2010).

This evidence of a common neuronal network was also found with functional studies (Gray et al., 2003). The conclusion that fluid intelligence and working memory involve a distributed fronto-parietal network is consistent with the hypothesis that there is a strong dependency between fluid intelligence and working memory (Kyllonen & Christal, 1990; Colom et al., 2004, 2005; Martinez et al., 2011).

The existence of this common neuronal network is highly relevant to the current study because some authors have suggested that it is necessary to make sure that relevant brain regions are engaged by the training regimen if some transfer effects are expected (Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008a). The training program based on the n-back was designed with the idea that fluid intelligence and short-term memory capacity share capacity limitations (Halford et al., 2007). This common constraint suggests that improvements in WMC tasks might have a direct influence over fluid intelligence. The study of Dahlin et al. (2008b), where young adults were trained on six

different updating tasks for five weeks, supported this idea. They found transfer to an untrained 3-back task, but not to an untrained Stroop task. The untrained 3-back task shared with training tasks activations in the striatum and updating process, while Stroop task did not. Therefore, they concluded that transfer could occur if the training and transfer task engage similar processes and brain regions.

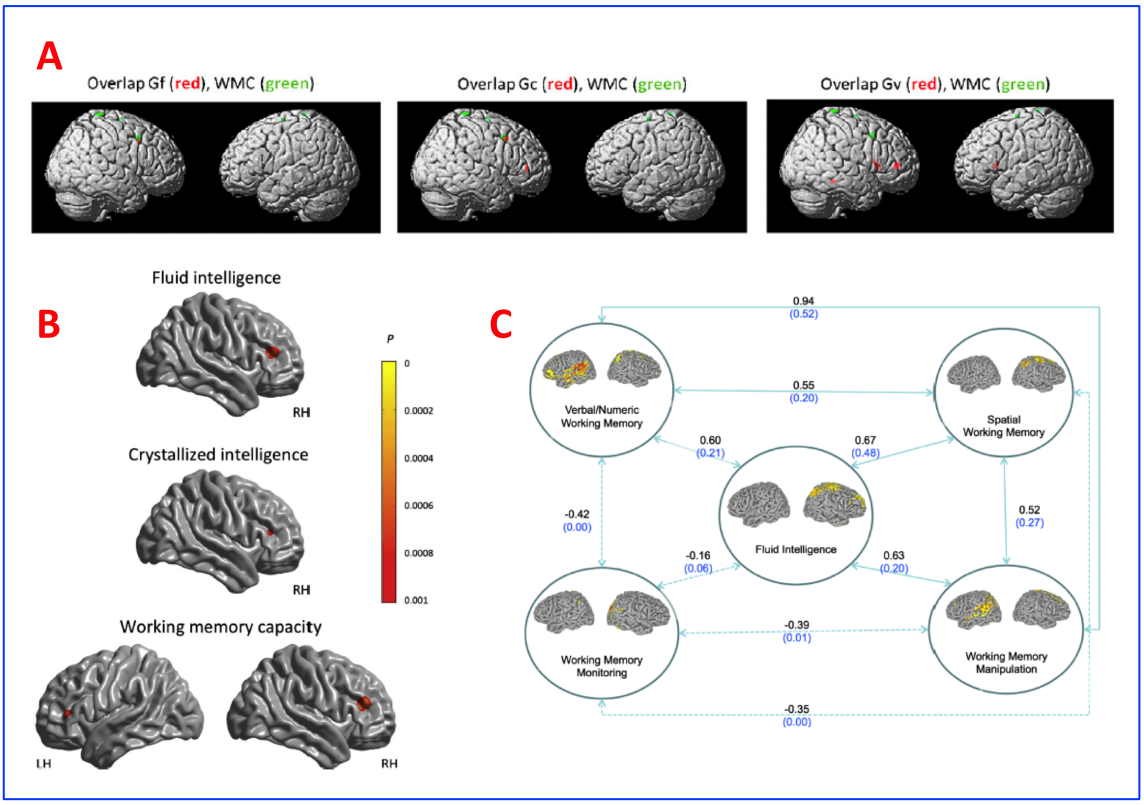


Figure 3.16. Clusters shared by intelligence and working memory factor defined at the latent level. Panels A and B show the results found by Colom et al., (2013b) applying VBM and SBM techniques respectively. Finally, panel C depicts the results found by Barbey et al., (2014a) with lesion patients. Gf = fluid intelligence, Gc = crystallized intelligence, Gv = visual-spatial intelligence and WMC = working memory capacity.

3. 7. Induced changes in the brain through experience: The effects of Training Programs.

3.7.1. First programs.

The first studies that addressed the idea that brain could be malleable by experience were the studies by Maguire et al. (2000) and Gaser and Schlaug (2003). Maguire et al. (2000) found that posterior hippocampi of taxi drivers were significantly larger compared to control subjects (this result was replicated in a posterior longitudinal study; Woollett & Maguire, 2011). Gaser and Schlaug (2003) discovered grey matter volume

differences in motor, auditory, and visual–spatial brain regions when professional musicians were compared with a matched group of amateur musicians and non-musicians. However, both studies used a transversal design, but only longitudinal studies could help to understand which came first, the structural differences or the experience (Buschkuehl, Jaeggi & Jonides, 2012; May, 2011).

Different programs were carried out and their effects were studied with behavioural and neuroimaging approaches. In behavioural level, the effectiveness of a training program can be measured in different ways: (a) changes in the trained task, (b) changes in an untrained task tapping the same cognitive dimension (near-transfer), and (c) changes in an untrained task tapping a different cognitive dimension (far-transfer). It is assumed that when the intervention has an effect on the criterion measure/s, the task used during training is thought to have the potential for inducing cognitive plasticity. Moreover, training-induced plasticity must affect a general mechanism that at least partially underlies the training and the criterion tasks. However, the mechanisms underlying training effects for cognitive skills are still largely unknown (Buschkuehl et al., 2012). The study of neuronal plasticity via neuroimaging methods is one promising way to explore such underlying mechanisms (Willis & Schaie, 2009), since the brain, like the body, can be trained (Hunt, 2012a). With this purpose in mind, different studies have been designed to study structural changes associated with diverse training programs in young people.

Draganski et al. (2004) designed the first longitudinal study applying a neuroimaging approach (VBM). They trained a group of 12 participants in a motor-task (Juggling) during 3 months. The participants of this study were scanned three times (before training, after training and three months after the end of the training), and their scans were compared with a non-trained group matched in age and sex. The longitudinal analysis showed bilateral expansion in grey matter (VBM approach) in the mid-temporal area and in the left posterior intraparietal sulcus. This expansion decreased in the third scan (see Figure 3.17).

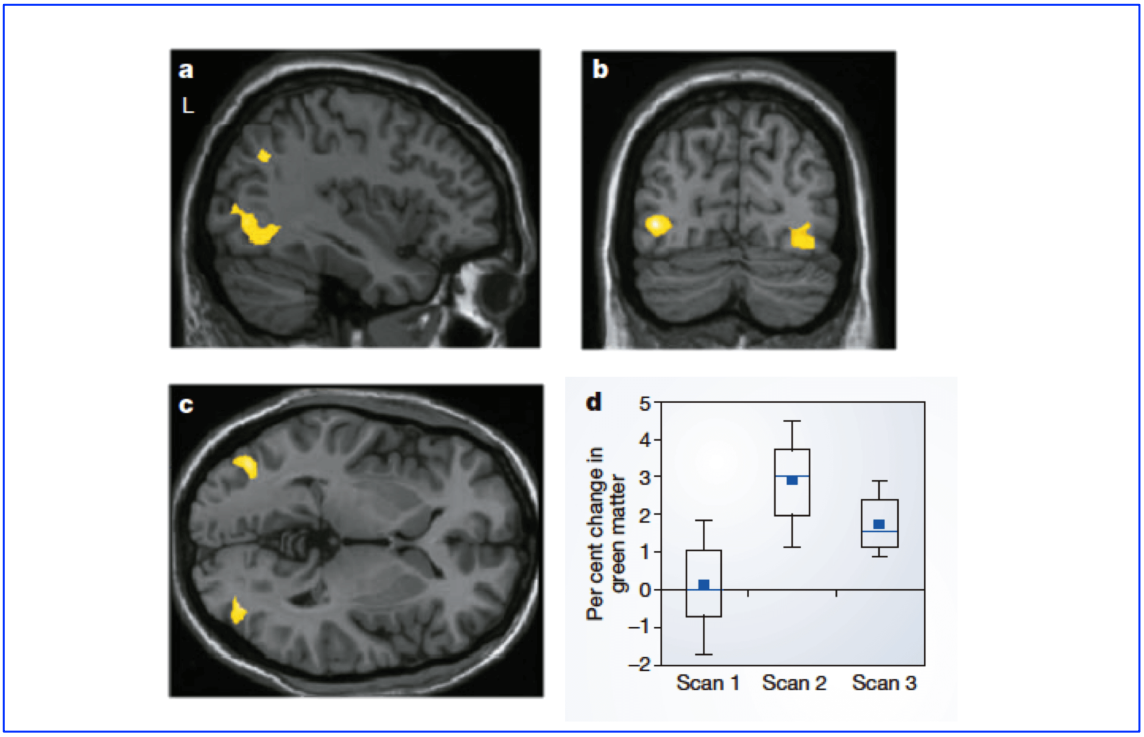


Figure 3.17. Changes in brain structure induced after learning to juggle. A–C, statistical parametric maps showing the areas with transient structural changes in grey matter for the jugglers group compared with non-juggler controls. D, relative grey-matter change in the peak voxel for all jugglers over the three time points. The box plot shows the standard deviation, range and the mean for each time point. Adapted from Draganski et al., (2004).

The effects of different training programs over brain structures have been also tested with neuroimaging methods in adults. These programs range from visual-motor tasks such as, juggling (for e.g., Draganski et al., 2004), golf (Bezzola, Mérillat, Gaser, & Jäncke, 2011) or balancing (Taubert et al., 2010) to cognitive tasks such as deciphering Morse code (Schmidt-Wilcke, Rosengarth, Luerding, Bogdahn, & Greenlee 2010), working memory (for e.g., Takeuchi et al., 2010), learning for an exam (Draganski et al., 2006), or playing videogames (e.g. Colom et al., 2012). See Buschkuehl et al., (2012) or Taya, Sun, Babiloni, Thakor, and Bezerianos, (2015) for a review of functional MRI studies.

3.7.2. *Cognitive programs.*

There are a wide variety of cognitive programs that have been employed to study the effects of short-term interventions over structural brain indices. Draganski et al. (2006) applied VBM techniques to detect possible structural brain changes associated with

learning in a group of medical students (see also Ceccarelli et al., 2009). They detected grey matter increases in the posterior and lateral parietal cortex bilaterally and in the posterior hippocampus. Also, these changes remained in the third scan (3 months after exam). In fact, the hippocampus showed a higher increment in the third scan. Similar effects were found in a short intervention (3 days) carried out by Kwok et al., (2011), which was dedicated to learning newly defined and named subcategories of the universal categories green and blue. The study by Mackey, Whitaker, and Bunge (2012) focused on changes of white matter properties after a 3 months reasoning training with students of law. They found decrements of radial diffusivity (RD) in white matter connecting frontal cortices, and of mean diffusivity (MD) within frontal and parietal lobe white matter. Moreover, participants exhibiting larger gains exhibited greater decreases in MD in the right internal capsule.

Ilg et al., (2008) investigated the effects of a 2-week practice period of mirror reading. They combined functional and structural brain data in order to investigate overlap between regions. Decrements in the activation of right superior parietal cortex and increments in right dorsal occipital cortex were found. The increment in activation was associated with an increment of grey matter volume in this area. A marginal overlap between functional and structural changes was found in occipito-temporal/inferior temporal cortex after training program dedicated to deciphering Morse code (Schmidt-Wilcke et al., 2010). However, Haier, Karama, Leyba and Jung (2009) found an increment of cortical thickness in frontal and temporal areas after playing an online version of *Tetris* during three months (1.5 h per week on average), but these increments did not overlap with decrements found in functional signal, mostly in the frontal lobe.

Colom et al., (2012) compared a group that played with *Professor Layton & The Pandora Box* (by Nintendo) – a cognitively complex commercial videogame – with a control group. The training was completed across four weeks (4 h per week). This study failed to find significant changes in cortical thickness and cortical surface area after completing the game, although VBM analyses showed volume changes in regions associated with intelligence. White matter showed increments in axial diffusivity at the hippocampal cingulum and inferior longitudinal fasciculus. Kühn et al., (2014) compared a control with a video gaming training group that was trained for 2 months for at least 30 min per day with a platformer game (*Super Mario 64*). Significant

increments were found in right hippocampal formation, right dorsolateral prefrontal cortex, and bilateral cerebellum in the training group.

To our knowledge, the study of Takeuchi et al. (2011) is the only investigating grey matter volume changes following working memory training. They reported reductions in the bilateral dorsolateral prefrontal cortex, right inferior parietal lobule, left paracentral lobule, and left superior temporal gyrus after completing a program based on an adaptive mental multiplication task (see Figure 3.18a). Regarding white matter changes, Takeuchi et al. (2010) studied the effect of working memory training over fractional anisotropy (FA). Participants completed 2 months of training with a working memory program based in three working memory tasks. They found an increase in FA in parietal as well as frontal cortices (see Figure 3.18b). However, they failed to include a control group, and, therefore, it was impossible to determine the causality of FA increases. Nevertheless, the number of completed training sessions correlated positively with pre- to post-training increases in FA, suggesting at least some neural effects as a consequence of the intervention.

Lövdén et al., (2010) trained young (20 – 31 years old) and old adults (65 - 80 years old) during 100 sessions in different tasks engaging: working memory (updating and shifting processes), episodic memory, and processing speed. These authors focused in changes of the corpus callosum. DTI-based analyses showed similar increases in white-matter microstructure for both training groups (younger and older). Changes were higher in the anterior part of the corpus callosum. However, no significant correlations were obtained between white matter changes and behavioral performance improvements.

3.7.3. Limitations.

Thomas and Baker (2013a,b) reviewed structural neuroimaging studies analyzing grey and white matter changes after behavioral training. They concluded that, given the microscopic nature of these changes in animals and the relatively low spatial resolution of MRI, it is unclear that such changes can be reliably detected in humans with current neuroimaging approaches. They proposed that future studies should analyze the interaction between groups (training vs. control) and times (before vs. after training). Also, straightforward hypotheses with respect to brain regions sensitive to training must

be stated in advance. These points are frequently ignored in current research (Thomas & Baker, 2013a,b).

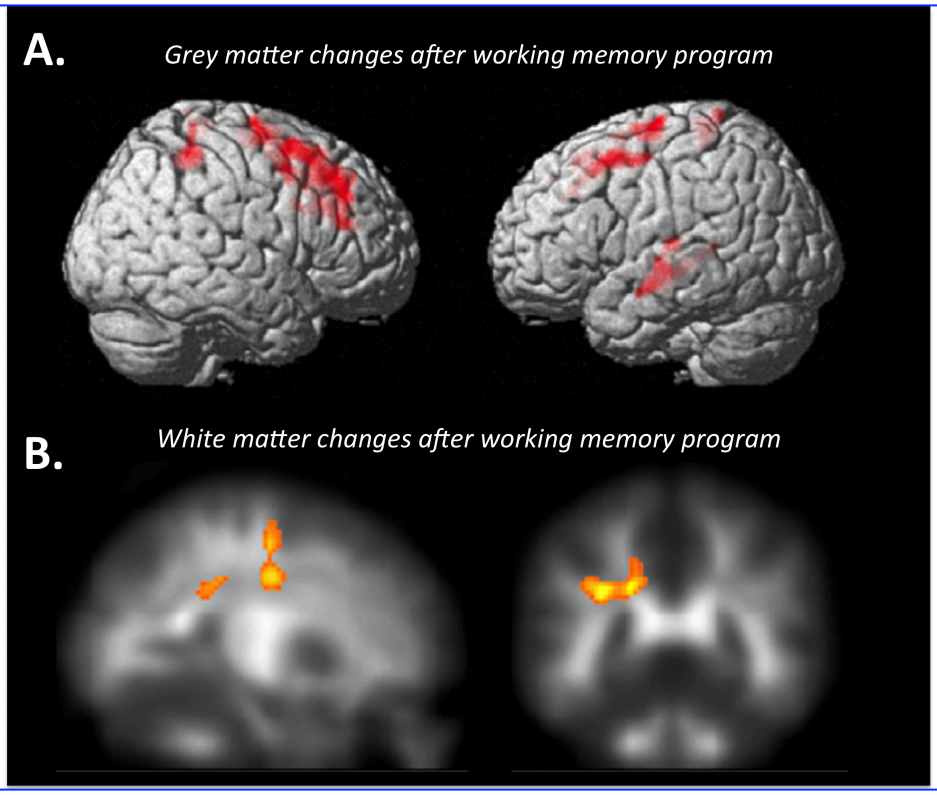


Figure 3.18. Top panel shows changes in grey matter found by Takeuchi et al., (2011) after an adaptive working memory program. Bottom panel shows significant changes in white matter after working memory training in the study carried out by Takeuchi et al., (2010).

Lövdén, Wenger, Mårtensson, Lindenberger, and Bäckman (2013) pointed that beyond design and statistical analyses, another weakness in this field is the heavy dependence on VBM to estimate structural changes in regional grey matter. VBM has important advantages, but also has important weaknesses since, for example, it is very sensitivity to small differences in preprocessing (Thomas & Baker, 2012a).

Furthermore, the exact cellular mechanisms underlying these variations, as well as their possible impact on the MR signal, are only weakly understood (Draganski & May, 2008). Buschkuehl et al. (2012) summarize different structural changes following training. Changes in grey matter signals could be indicative of dendritic spine growth leading to the sprouting of new connections and/or a modification of existing connections. Nevertheless, other mechanisms, such as neurogenesis, genesis of glial cells, increase of cell sizes, or angiogenesis, may be present. However, the latter

mechanisms are presumably slower than dendritic spine growth and associated synaptogenesis.

Changes in FA can be produced by changes in myelination, resulting in an increased conduction velocity, and signal synchronization, which might facilitate the interaction among neural circuits. This can, in turn, regulate observed performance improvements. Other potential mechanisms could be a change in density and/or in the diameter of axons. Finally, as glial cells seem to play an important role in inducing structural changes in grey and white matter structures, there is the possibility for a common primary mechanism driving structural changes and, as such, learning-related plasticity (Scholz et al., 2009).

Zatorre, Fields, and Johansen-Berg (2012) noted that behaviorally evoked changes in the brain might involve increment or decrement in the considered biological signals, since the interaction between structural and functional changes in the brain is complex. See Figure 3.19 with examples of these processes.

One final relevant question is: how long must the training program be to evoke macroscopic structural changes in the human brain? There are studies suggesting that at least three months are required by stem cells to differentiate into neurons. However, there are intervention studies showing changes in grey matter in the very short-term (Sagi et al., 2012). Myelination can be changed by electrical activity in two days (Ishibashi et al., 2006). According to this results, the time interval considered in our study seems appropriate for maximizing the probability of evoking detectable structural changes in the brain.

3. 8. Conclusions.

The brain changes spontaneously across lifespan but it also responds to short-term cognitive interventions. A review of previous literature shows that grey matter indices show a curve with U inverted form (Raznahan et al., 2011). However, while the decrement of cortical thickness during adolescence and adulthood is relatively replicated, (e.g. Raznahan et al., 2011; Shaw et al., 2006, 2008) the results for cortical surface area are unclear since increment and decrement were found in several independent studies (e.g., Raznahan et al., 2011; Schnack et al., 2014; Zhou et al.,

2015). This is highly relevant to our study, since we hypothesized that changes in the control group should be consistent with the changes associated to age. Respecting white matter, we theorized an increment in white matter since the decrement is usually found over 40/50 years old (Cassey et al., 2005; Giedd et al., 1999; Giedd & Rapoport, 2010; Sowell et al. 2003).

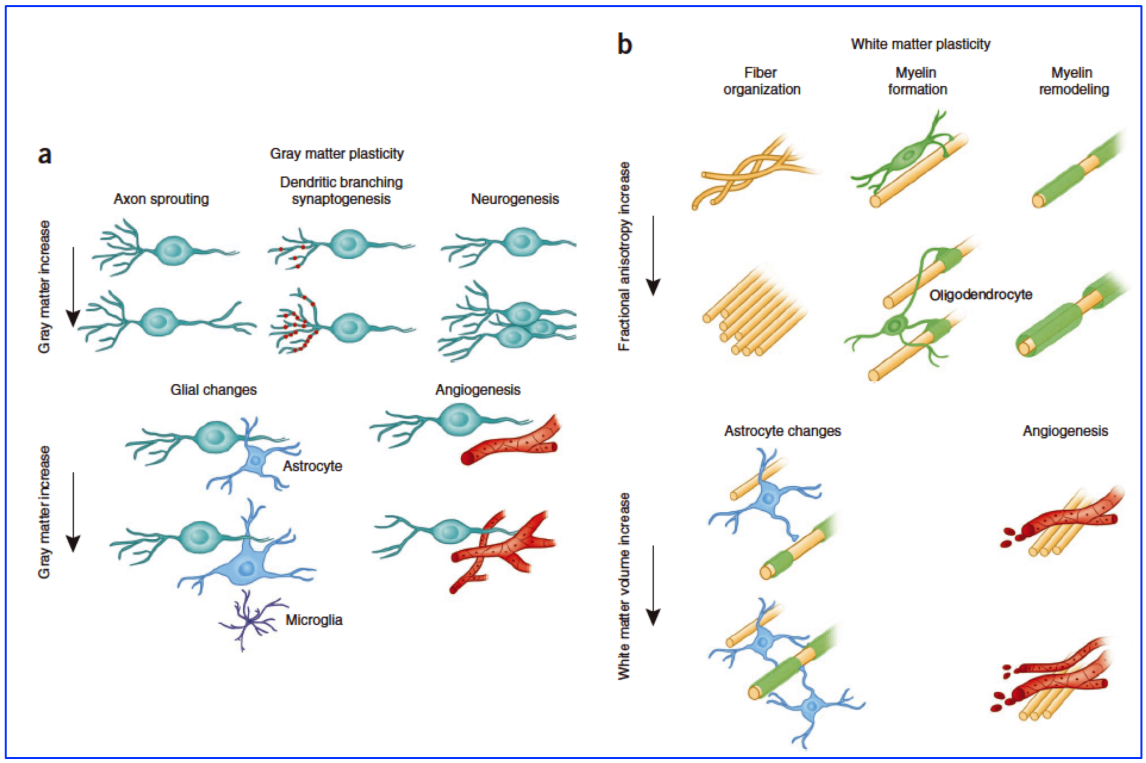


Figure 3.19. Candidate cellular mechanisms. (a) Cellular events in grey matter regions underlying changes detected by MRI during learning include axon sprouting, dendritic branching and synaptogenesis, neurogenesis, changes in glial number and morphology, and angiogenesis. (b) Changes in white matter regions include alterations in fiber organization, which could include axon branching, sprouting, packing density, axon diameter, fiber crossing and the number of axons; myelination of unmyelinated axons; changes in myelin thickness and morphology; changes in astrocyte morphology or number; and angiogenesis. Adapted from Zatorre et al., (2012).

On the other hand, we considered that changes in the training group should be associated to cognitive training program (Dahlin et al., 2008b). Our training program is based on a working memory task (n-back) that was found to effectively change the scores in fluid intelligence (e.g. Jaeggi et al., 2008). Therefore, fronto-parietal networks were the candidates expected to be more sensitive to changes, since several studies found overlap between fluid intelligence and working memory in these regions with structural (Barbey et al., 2014a; Colom et al., 2013b) and functional approaches (Gray

et al., 2003). However, available studies show important weaknesses from a methodological (Thomas & Baker, 2013a) and conceptual perspective (Zatorre et al., 2012).

Lövdén et al. (2013) considered that Surface-based morphometry (SBM) methods have important advantages over traditional Voxel-based morphometry (VBM) methods. SBM allows the study of independent grey matter indices (cortical thickness and cortical surface area), showing a different genetic substrate (Chen et al., 2013). White matter studies are generally focused on fractional anisotropy (FA), but there is a lack of studies considering whole brain white matter organization. In this respect, Taya et al. (2015) considered that the connectome approach could be applied to studying alterations of the brain network induced by cognitive training.

Only three structural studies have considered working memory programs for training young people (Lövdén et al., 2010; Takeuchi et al., 2010, 2011). Results suggested that grey and white matter in fronto-parietal and middle temporal regions were sensitive to training (Takeuchi et al., 2011). Lövdén et al. (2013) found changes in the anterior part of the corpus callosum, which links parietal and frontal regions. Therefore, frontal and parietal regions seem sensitive to cognitive training based in working memory tasks. Nevertheless, structural changes associated with a cognitive program based on the n-back task are still unknown.

In the recent research reported by Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides (2014), participants were trained during 7 days with the n-back task (visual or auditory modality). Greater neural activation in the contrast 4-back minus 1-back was found. The largest effect was observed in bilateral BA 6 (frontal) and in one cluster in BA 19 (occipital). Analyzing resting state changes, increases after training were found in frontal (left BA 6) and parietal regions (left BA 39); decrements were observed in right parietal BA 5. Furthermore, blood flow at rest in the trained brain regions increased along with task proficiency. Conjunction analysis revealed that left BA 6 was a common brain area.

In closing, structural studies considering cognitive training programs based on the n-back task are greatly required. The available behavioral evidence suggests that this sort of training may have some impact on fluid reasoning improvements (Au et al., 2014).

The research to be reported in the present work was designed to address this issue as exhaustively as possible.

CHAPTER 4: Framework and Research Hypothesis.

4. 1. Introduction.

This chapter describes the main framework of the present research, along with the key hypothesis. Research questions are asked at the two levels of analysis: psychological and biological. Details regarding the follow-up study will also be provided.

4. 2. Psychological level.

4.2.1. *Cognitive Training program*

The program employed in this work is based on the adaptive n-back task (Jaeggi et al., 2008). This task was chosen because there is evidence showing a positive impact on fluid reasoning (Jaeggi et al., 2008, 2010a, 2011; Kundu et al., 2013; Schweizer et al., 2011; Stephenson & Halpern, 2013; Vartanian et al., 2013). Fluid intelligence is crucial for many relevant social outcomes (Deary, 2012; Deary et al., 2010; Gottfredson, 1997; Nisbett et al., 2012). The meta-analysis by Au et al. (2014)[1] concluded that an increment of 3.6 IQ points was found after n-back training in young adults.

Nevertheless, in the current work, some of Jaeggi's guidelines (Jaeggi et al., 2008) were modified. Specifically, our training program started with 8 sessions based on the single n-back task (4 visual and 4 auditory). After these adaptive training sessions, 16 sessions with the dual n-back were completed. Furthermore, the training sessions were completed across twelve weeks (3 months), twice per week (following recommendations by Klingberg, 2010). The n-back task is complex and participants found it difficult to follow the provided instructions. Thus, it was assumed that practicing with a single version would be beneficial in this regard. Also, note that Jaeggi et al., (2010) found that practice with the single n-back task is also related with improvements in fluid reasoning.

Susanne M. Jaeggi suggested that real engagement is crucial for satisfactory performance during the n-back task and to maximize transfer effects to untrained tasks (near and far transfer effects). Therefore, feedback about their performance at the end of each session was added. Also, a comparative graph between their performance and the

[1] Part of the behavioral analysis of this work is included in this meta-analysis (Colom et al., 2013a).

performance of a hypothetical sample of American students was shown to all participants at the midpoint of the program. These strategies (feedback and external comparison) were used to maintain the engagement of our participants, which was measured with a questionnaire (Tapia & de la Red Fadrique, 2007).

4.2.2. *Psychological measures*

While performance on the n-back task has been related with fluid intelligence (Jaeggi et al., 2010a, 2010b), this task is not considered a straightforward measure of working memory capacity (Jaeggi et al., 2010b). Here, we were interested in checking whether n-back performance is related to both fluid intelligence and working memory. Also, crystallized intelligence (Gc) and attention control (ATT) were measured. Both constructs have been related to Gf and WMC in previous studies (see Martínez et al., 2011). We also administered a set of personality traits (neuroticism, extraversion, openness to experience, agreeableness and conscientiousness). Finally, we analyzed the effect of the level achieved in the training regime creating two groups of participants showing standard and top performance. This strategy was employed in previous reports (e.g., Jaeggi et al., 2011). Also, this approach has the benefit of rendering the results readily interpretable, allowing comparisons between top and standard performers. However, the method loses power by categorizing a continuous variable.

4.2.3. *Transfer effects*

As discussed before, Jaeggi et al. (2008) assumed that intelligence and working memory share cognitive processes and this fact may stimulate expected transfer effects (Halford et al., 2007). These improvements in the adaptive cognitive training program are expected to promote significant increments in fluid intelligence mainly because their common limitations (based on the reliable temporary storage of the relevant information) will be boosted. Here we administered diverse intelligence and cognitive measures (three measures for each psychological factor) before and after completing the training program to go beyond specific tests and tasks.

4. 3. Biological level.

4.3.1. *Cortical thickness (CT) and cortical surface area (CSA)*

CT and CSA indices were computed with CIVET 2.0 pipeline developed by Montreal Neurological Institute (MNI; Ad-Dab'bagh et al., 2006)[2]. Two levels of analysis were carried out: (a) vertex level and (b) region of interest level (ROIs). In the latter instance, twelve regions of interest were defined for CT, as well as for CSA. The ROIs were distinguished by their genetic substrate (Chen et al., 2012, 2013). The Chen et al.'s (2012, 2013) studies applied a data-driven approach for generating a genetically based parcellation for CSA and CT. As a result, they found an anterior-posterior division for CSA, along with a dorsal-ventral division for CT. The genetic patterning of both gray matter indices corresponds to functional specializations and their genetic contributions show very small correlations. Further, principles underlying genetically defined regions are different for CSA and CT: the regions/clusters for CSA show large genetic proximity within the same brain lobe, whereas regions for CT show remarkable genetic relatedness regarding maturational timing (primary vs. association cortex). Therefore, CT and CSA show differential neurodevelopmental mechanisms. Our study focused on the analysis of which brain regions, under remarkable genetic control, are sensitive to the completed challenging cognitive training.

The same analytic strategy was followed in both cases: correlations of these indices with n-back performance and a standardized measure of change (Jaeggi et al., 2011) were computed to analyze the interaction (Group x Time). These analyses were replicated with the division according to n-back performance (top and standard performers)

4.3.2. *Structural connectivity*

Here we used a global tractography algorithm based on graph theory (Iturria-Medina et al., 2007) to compute connectivity matrices. The specific details of the algorithm and the process of reconstruction are discussed in the empirical section. Graph theory allows studying the connectome, which is especially relevant to analyze the development of brain connectivity and plasticity (Tymofiyeva et al., 2014). Also, recent reviews suggest

[2] Analyses were also computed with a previous version of CIVET 1.12 and the results were similar.

that the connectome is an interesting framework for these sorts of analyses (Taya et al., 2015).

We studied changes (a) in connectivity at the node level, (b) in graph-theory indices (clustering and path length) and (c) at the network level applying Network-based statistic (NBS). Also, analyses were computed for group comparisons and dividing the training group according to their performance (top vs. standard).

4. 4. Follow-up

We also explored whether improvements observed during the training program were maintained one year after ending the program. For that purpose, a new version of the n-back was administered (Román et al., 2015).

Academic performance 3 months, two years and three years after ending the training program was also measured. To our knowledge, this will be the first analysis checking probable social correlates of the training (Buschkuehl, & Jaeggi, 2010; Estrada et al., 2015; Hunt, 2012a; Jensen, 1998).

4. 5. Research questions and hypotheses

These are the main research questions (Q) and hypotheses:

Q1: How was performance across the training program?

This question is addressed in chapter 6. Correlations between n-back performance and different psychological traits are computed and these are the hypotheses:

(a) n-back performance will be related with fluid intelligence (Gf) and working memory capacity (WMC). Previous studies have shown that n-back performance might measure individual differences in Gf and requires working memory related processes (Jaeggi et al., 2010b).

(b) Neuroticism (N) and conscientiousness (C) will be related with n-back performance differences. Specifically, low N and high C will predict greater performance. Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, (2012) showed that these personality traits are related with transfer to untrained measures.

We also studied transfer to Gf, Gc, WMC and ATT. Significant effects are expected for working memory (near-transfer) and fluid intelligence (far-transfer). Changes in crystallized intelligence and attention control are not expected. Fluid intelligence and working memory require the reliable preservation of relevant information in the short-term (Carpenter et al., 1990), whereas this is not the case for crystallized intelligence and attention control. Gc requires the recovery of the relevant information from long-term memory and attention control does not require any short-term storage. Finally, differences between top and standard performers are calculated to test if there were differences in transfer to untrained tasks. As noted above, we expect greater transfer effects for participants showing the largest improvements in the training program (top performers) (Jaeggi et al., 2011).

Q2: Are there structural changes in the brain attributable to the training program?

This question is addressed in chapter 7 (focused on the changes in cortical thickness and cortical surface area) and chapter 8 (devoted to changes in structural connectivity).

The main prediction regarding potential gray matter responsiveness (chapter 7) after training includes anterior frontal, parietal, and middle temporal regions (see Takeuchi et al., 2011). Increases (or preservation) in these regions are expected because: (a) performance in the n-back task is related with intelligence and working memory, and parietal and anterior frontal regions are relevant for intelligence (Jung & Haier, 2007) as well as for working memory (Barbey et al., 2014a; Colom et al., 2008; Cowan, 2005), (b) cortical thickness preservation in anterior frontal and temporal areas has been found in previous training research focusing on visual-spatial skills (Haier et al., 2009; see also Bastenet al., 2015), and (c) they are known to be involved in information integration and evaluation cognitive processes, which are relevant for working memory (Buschman, Siegel, Roy, & Miller 2011; Hampson, Driesen, Skudlarski, Gore, & Constable, 2006). Furthermore, CT shows a spontaneous decrease with age (Wierenga et al., 2014; Zou et al., 2013). Therefore, we expect thinning for control group. For cortical surface area (CSA), changes during brain development have been less studied and results are less clear (see Burgaleta et al., 2014a; Wierenga et al., 2014; Zhou et al., 2015). Thus, no specific hypothesis for this cortical index are stated.

Note that Tang et al., (2010) –among others-- recommended analyzing neuroimaging data separately for males and females. To avoid issues related with the sex variable, only women are considered in the present study.

Regarding changes in structural connectivity (chapter 8), frontal and parietal regions were expected to be more sensitive to the training program (Takeuchi et al., 2010; Lövdén et al., 2010). Three analytic strategies are applied: changes (a) in connectivity at the node level, (b) in graph-theory indices and (c) at the network level applying Network-based statistic (NBS). The first approach is mainly exploratory, because there is a lack of previous results. Regarding graph-theory indices, developmental studies have shown decrements in early adolescence (Collin, & van den Heuvel, 2013; see Table 3.1). Therefore, it is hypothesized that the training program will evoke significant increments in clustering and decrements in path length, since we consider that the small-world properties of the network will increase. Finally, with respect to NBS, it is predicted that connectivity will increase in the network involved in the training program. This network may include bilateral frontal, bilateral parietal and left middle temporal regions (Yeo et al., 2014).

Q3: Are performance levels achieved in the training program maintained one year later?

Q4: Are there differences between training and control group in their academic performance?

Chapter 9 presents the follow-up results (Q3 and Q4). Near and far transfer effects are explored. To measure near-transfer, we applied a new version of the n-back task using neutral faces as stimuli (Román et al., 2015). Far-transfer is addressed measuring academic performance.

The prediction is that the achieved performance level will be maintained one year later (Melby-Lervåg & Hulme, 2013). To our knowledge, there are no studies that addressed the topic of transfer to academic performance. However, we expect higher scores in academic performance for the training group.

PART II: Research

The brain is wider than the sky.

- Emily Dickinson (1960)

The complete Poems of Emily Dickinson, 312

Here we present the follow-up results derived from the adaptive training program. There are three main questions (Q):

Q1: How was the performance across the training program?

Q2: Are there structural changes in the brain attributable to the training program?

Q3: Are performance levels achieved in the training program maintained one year later?

Q4: Are there academic performance differences related with individual differences in training success?

The answers to these questions will be addressed in five chapters comprising methods, behavioral results, neuroimaging findings, and general discussion.

Chapter five includes the basic design for the training program. Each stage and their results are presented. Chapter six is aimed at answering Q1 by analyzing (a) the improvements along the training program in the n-back performance, as well as the influence of baseline cognitive variability in the level achieved at the end of each session; (b) the transfer to the cognitive factors considered; and (c) the difference between average and top performers during the training program in both the baseline cognitive capacities and transfer effects. Chapter seven (Surface-Based Morphometry, SBM) and eight (Structural Connectivity, SC) include the neuroimaging results to answer Q2: (a) correlation of n-back performance with structural indices, (b) analyses of interaction (group x time) to detect regions where a group obtained higher changes than

the other group. Chapter nine is devoted to address Q3 and Q4 by analyzing (a) the behavioral results in a parallel n-back task one year after finishing the training program (follow-up); and (b) the impact of cognitive training over academic grades. The last chapter comprises the general discussion along with limitations and future research.

CHAPTER 5: Methodology and Basic Design.

5. 1. Introduction.

This was a pretest – posttest study. First, the psychological assessment and MRI scan protocols for time 1 (before training) are described. Secondly, the n-back training program is detailed. Thirdly, the psychological assessment and MRI scan protocols in time 2 (after training) are specified. Finally, the task administered at the follow-up is presented. Figure 5.1 shows the timeline.

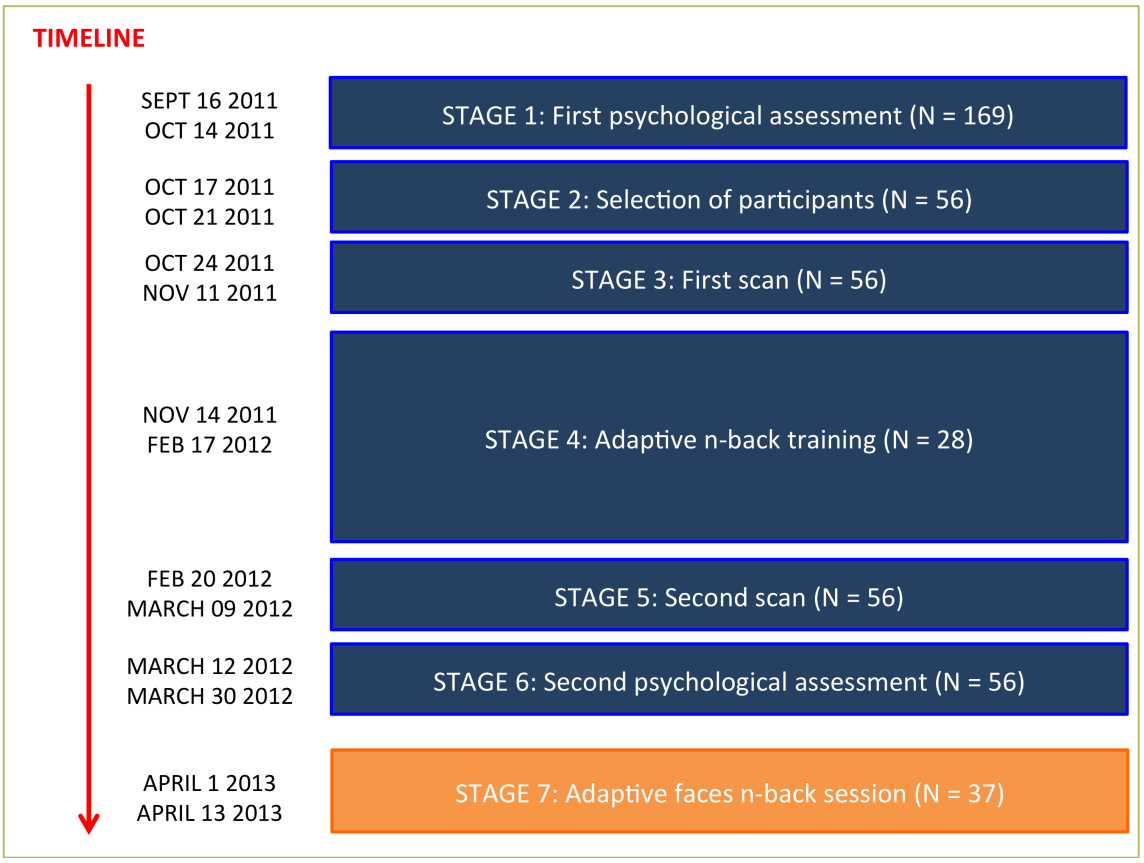


Figure 5.1. Timeline of our research.

5. 2. Stage 1: First psychological assessment.

5. 2. 1. *Participants.*

One hundred and sixty nine university undergraduates (128 females) aged between 17 and 49 years (mean 18.90, standard deviation 3.71) completed a cognitive battery, comprising 12 tasks assessing fluid-abstract intelligence (Gf), crystallized-verbal intelligence (Gc), working memory capacity (WMC), and attention control (ATT).

5. 2. 2. *Procedure.*

Three sessions were completed. Each session lasted approx. one hour and participants were assessed in groups of no more than 30 students. The first session assessed fluid and crystallized intelligence, whereas in the second session, participants completed the cognitive tasks (working memory capacity and attention). The personality inventory was completed on-line.

5. 2. 3. *Cognitive Battery and Personality Inventory*

Fluid ability (Gf) was measured using the Raven Advanced Progressive Matrices Test (RAPM; Raven, Raven & Court, 2004), the abstract reasoning subtests from the Differential Aptitude Test (DAT-AR) battery (Bennett, Seashore, & Wesman, 2005), and the inductive reasoning subtests from the Primary Mental Abilities (PMA-R) battery (Thurstone & Thurstone, 1968). The RAPM (Time = 18 minutes) comprises a matrix figure with three rows and three columns. Among eight possible alternatives, the one completing the 3 x 3 matrix figure must be chosen (max. score = 18). DAT-AR (Time = 10 minutes) is a series test based on abstract figures. Successive figures follow a given rule, so the one continuing the series must be chosen from several alternatives (max. score = 20). PMA-R (Time = 3 minutes) comprises letter series items. The rule (or rules) underlying a given sequence must be extracted to select the correct alternative (max. score = 15). See Figure 5.2 for an example of Gf tests.

Crystallized ability (Gc) was measured using the verbal and numerical reasoning subtests from the DAT (VR and NR), along with the vocabulary subtests from the PMA (V). All tests were administered in their Spanish version. DAT-VR (Time = 10 minutes) is based on sentences stated like an analogy. The first and last words from the sentence are missing, and a pair of words completing the sentence must be selected. The screening version comprising odd items only was administered (max. score = 20). PMA-V (Time = 2 minutes) is a synonym test based on the meaning of words that must be evaluated against a given model word (max. score = 25). DAT-NR (Time = 10 minutes) consists of quantitative reasoning problems. The screening version comprising odd items only was administered (max. score = 20). See Figure 5.2 for an example of Gc tests.

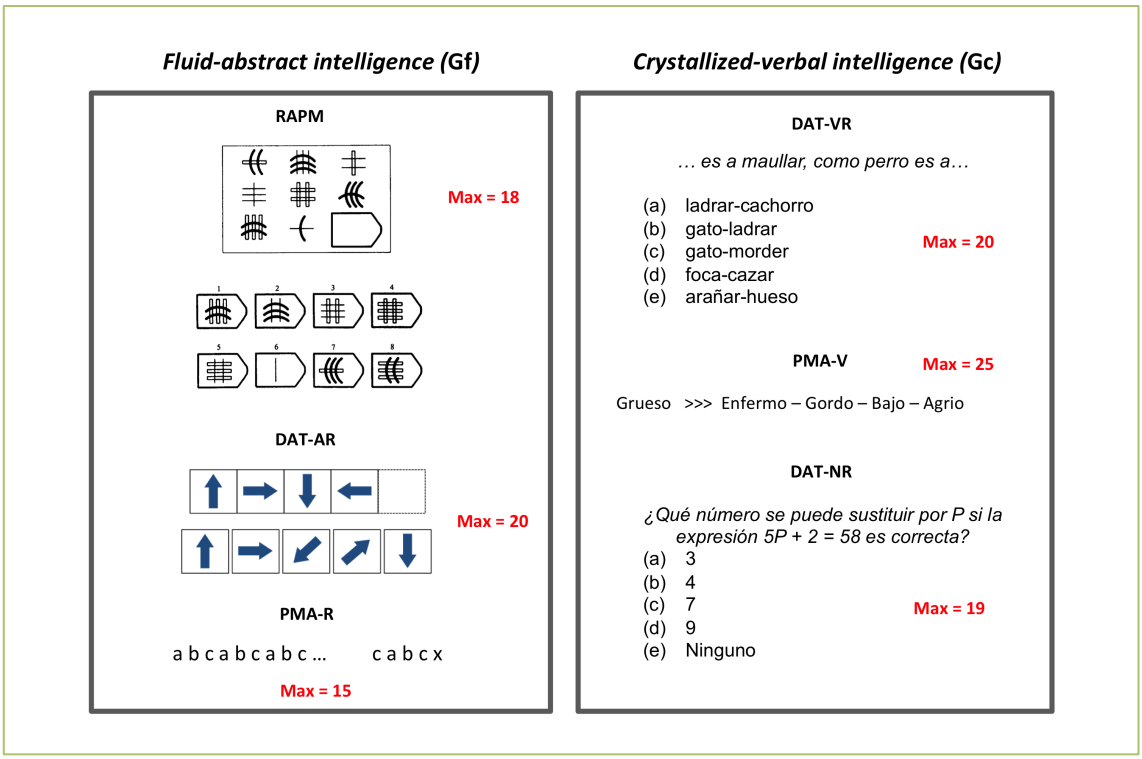


Figure 5.2. Examples of tests administered for fluid-abstract intelligence (left) and crystallized verbal intelligence (right).

The computation span task includes a verification task and a recall task. Six seconds are allowed to see the math equation without a time limit for verifying its accuracy. The displayed solution, irrespective of its accuracy, must be serially remembered at the end of a given set. Each math equation includes two operations using digits from 1 to 10. The solutions are always single-digit numbers. Trials range from three to seven equation/solutions (5 levels x 3 trials each = 15 trials total). In the dot matrix task, a matrix equation must be verified and a dot location, displayed in a five x five grid, must be retained. The matrix equation is presented during a maximum of 4.5 seconds to add or subtract simple line drawings. Once the response is given, the grid comprising the to-be remembered dot is displayed for 1.5 seconds. After a given set of equation-grid pairs, the grid spaces that contained dots must be recalled clicking with the mouse on an empty grid. Trials increase in size from two to five equations and dots (4 levels x 3 trials = 12 trials total). The score for these three WMC tasks is the number of hits in the verification and recalling tasks. Figure 5.3 depicts an example of items of WMC.

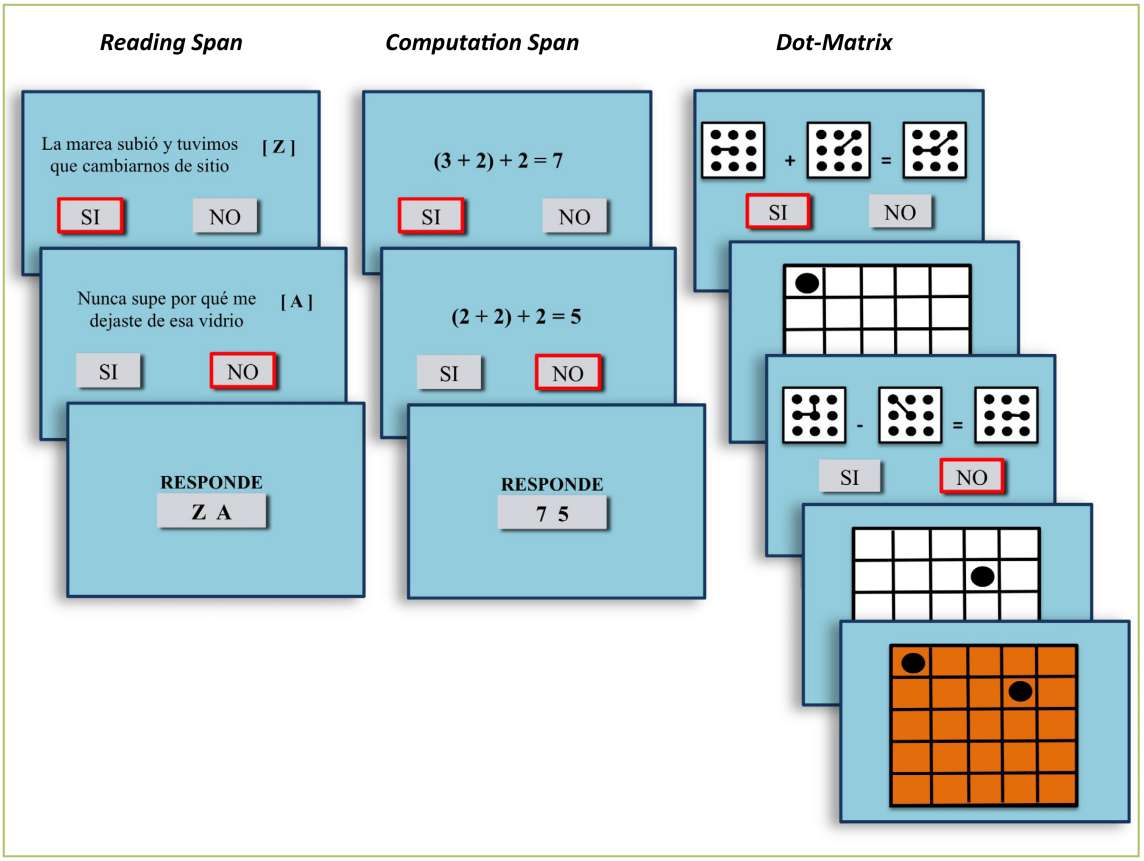


Figure 5.3. Examples of items from the task administered to measure working memory capacity.

Attention control (ATT) was measured using verbal and numerical versions of the flanker task, along with a spatial variant of the Simon task (Colom et al., 2010). The verbal and quantitative tasks require deciding, as fast as possible, if the letter/digit presented in the center of a set of three letters/digits is vowel/odd or consonant/even. The target (e.g. vowel/odd) can be surrounded by compatible (e.g. vowel/odd) or incompatible (e.g. consonant/even) letters/digits. The spatial task (Simon task) requires deciding if an arrow (horizontally depicted) points to the left or to the right of a fixation point. The target arrow pointing to a given direction (e.g. to the left) can be presented at the left (e.g. compatible) or at the right (e.g. incompatible) of the fixation point. There are a total of 32 practice trials and 80 experimental trials. Half of the trials are compatible and they are randomly presented across the entire session. The mean reaction time for the incompatible trials is the dependent measure. Figure 5.4 shows an example of compatible and incompatible trials in ATT tasks.

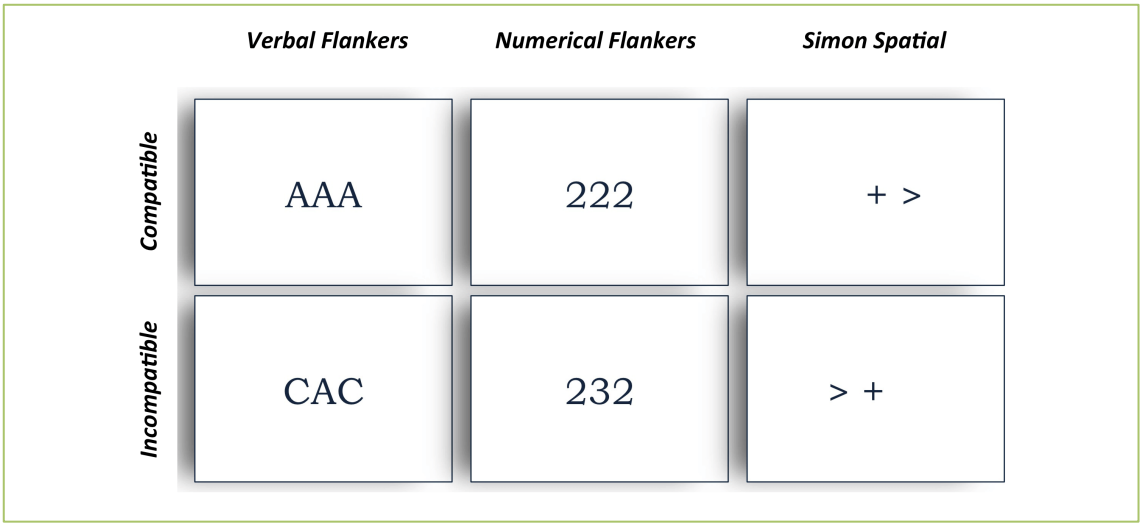


Figure 5.4. Example of compatible and incompatible trials in the control attention tasks.

Personality traits were measured using the Spanish version of the NEO-FFI questionnaire (Costa & McCrae, 1992). NEO-FFI is a 60-item questionnaire where five personality factors are computed: extraversion (E), agreeableness (A), conscientiousness (C), neuroticism (N), and openness to experience (O). The items are answered on a five point scale, ranging from «strongly disagree» to «strongly agree».

5. 2. 4. *Statistical analysis.*

Descriptive statistics were computed for the cognitive measures (Gf, Gc, WMC and ATT). For ATT, the compatibility effect (CE) was obtained by subtracting the reaction time for the compatible trials from the reaction time for the incompatible trials. This score is considered as a measure of the control of automatic responses (inhibition). Gf, Gc, WMC and ATT scores were submitted to confirmatory factor analysis (CFA). Confirmatory factor analyses were computed with MPLUS (Muthén & Muthén, 1998-2008) and their simultaneous latent relationships were analyzed. This model was computed twice: (a) with the complete sample and (b) for women only. A robust method of Maximum-Likelihood (MLM, Satorra & Bentler, 1994) was employed as the method of estimation. The fit of the model was assessed using three indices:

- **CMIN/DF (Chi Square/Degrees of Freedom)** ratio is considered given that it is a classic measure of model fit (Jöreskog, 1993). Values showing a good fit must be around 3.0 or lower.

- **RMSEA (Root Mean Square Error of Approximation)** index is sensitive to misspecification of the model. Values between 0 and 0.05 indicate a very good fit, values between 0.05 and 0.08 indicate a reasonable fit, and values greater than .10 indicate a poor fit (Byrne, 1998).

- **CFI (Comparative Fit Index)** (Bentler, 1990) is one of the measures least affected by sample size (Fan, Thompson & Wang, 1999). This statistic ranges between 0.0 and 1.0 with values closer to 1.0 indicating a good fit. A cut-off criterion of CFI ≥ 0.90 is presently recognized as a reasonable fit, CFI ≥ 0.95 indicates a good fit (Hu & Bentler, 1999).

5. 2. 5. *Results*.

The descriptive statistics for the intelligence tests and cognitive tasks for all participants (N = 169) are reported in Supplementary Table 1. CFA results are shown in Figure 5.5.

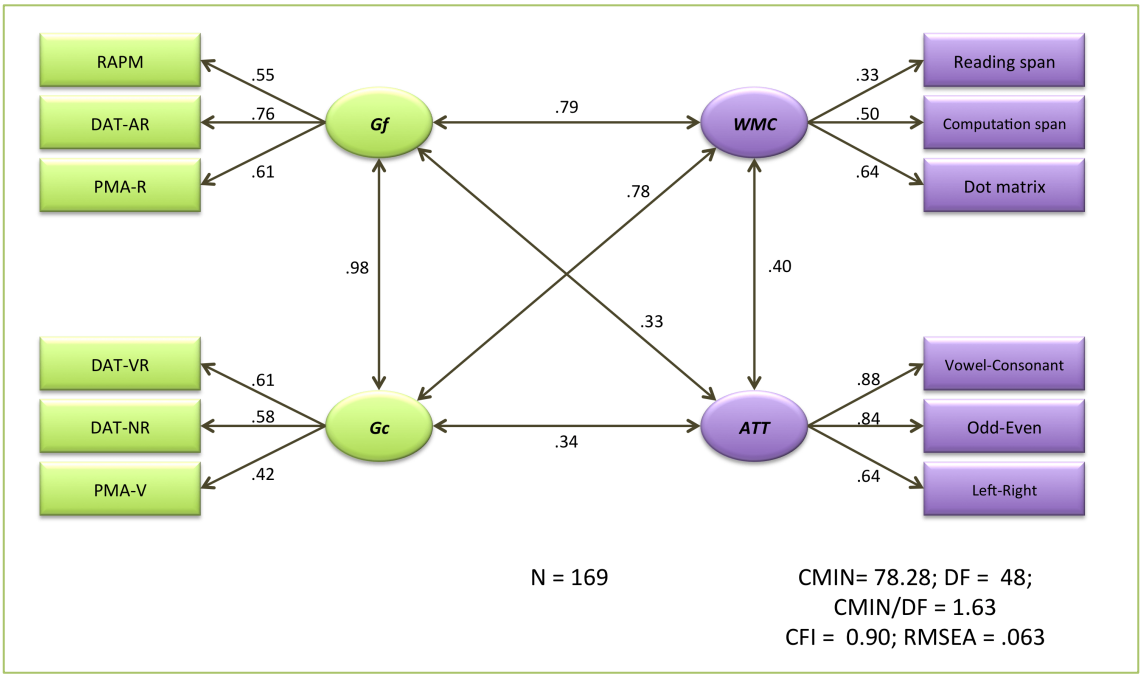


Figure 5.5. Confirmatory factor analysis (N = 169). Gf = abstract/fluid intelligence, Gc = verbal/crystallized intelligence, WMC = working memory capacity, ATT = attention control. Green = intelligence measures and violet = cognitive processes.

The model fit was acceptable (CFI = 0.90; RMSEA = .063). The highest correlation (r = .98) was found between Gf and Gc. The correlation of WMC with intelligence factors

was higher than .70. Finally, ATT was the psychological construct less related with the remaining factors (*rs* from .33 to .40).

The same model was computed for women only and the obtained results are shown in Figure 5.6.

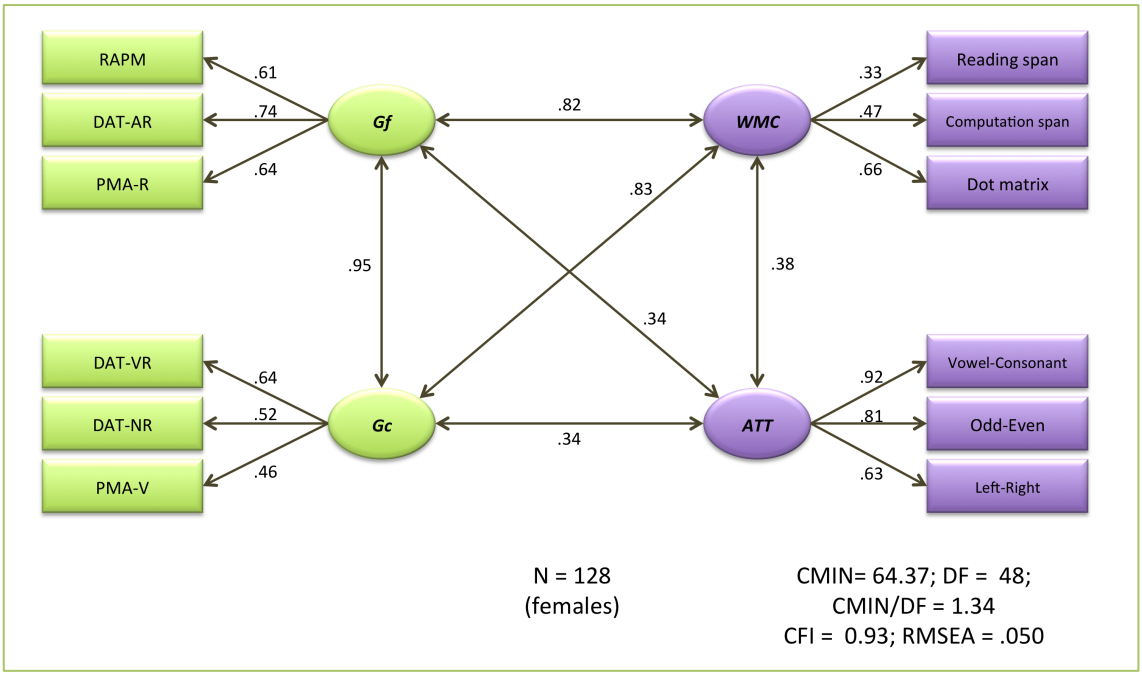


Figure 5.6. Confirmatory factor analysis (N = 128, women only). Gf = abstract/fluid intelligence, Gc = verbal/crystallized intelligence, WMC = working memory capacity, ATT = attention control. Green = intelligence measures and violet = cognitive processes.

This model showed a good fit (CFI = .93, RMSEA = .050). Again, the highest correlation was found between Gf and Gc (*r* = .95). WMC was also highly correlated with the intelligence factors (*r* > .80). Finally, ATT showed the same correlations with the remaining factors (*r* from .34 to .38).

5. 3. Stage 2: Recruitment.

5. 3. 1. *Pre-selection.*

Firstly, only women from the original sample (N = 128) were invited to an informative session because we were interested in controlling for sex differences in brain structure (Escorial et al., 2015; Ruigrok et al., 2014).

Secondly, volunteers attending the session (N = 100) were evaluated by the Edinburgh test for Spain (Oldfield, 1971). Only right-handed women were selected.

Thirdly, the volunteers completed a set of questions on medical or psychiatric disorders, as well as substance intake (see Appendix 1), along with a questionnaire regarding Magnetic Resonance Imaging (MRI; see Appendix 2).

5. 3. 2. *Final selection.*

Selection was performed according to the following guidelines: (a) women only, as noted above (b) right-handed, (c) age (range 18-25 years), (d) no tattoos close to the head, (e) no history of psychiatric disorders or addictions and (f) no chronic use of drugs.

Furthermore, participants were selected to represent a wide range of intelligence scores (mean = 100, standard deviation = 15; these scores are relative to the group assessed). The distribution for the original sample (N = 169) is shown in Figure 5.7.

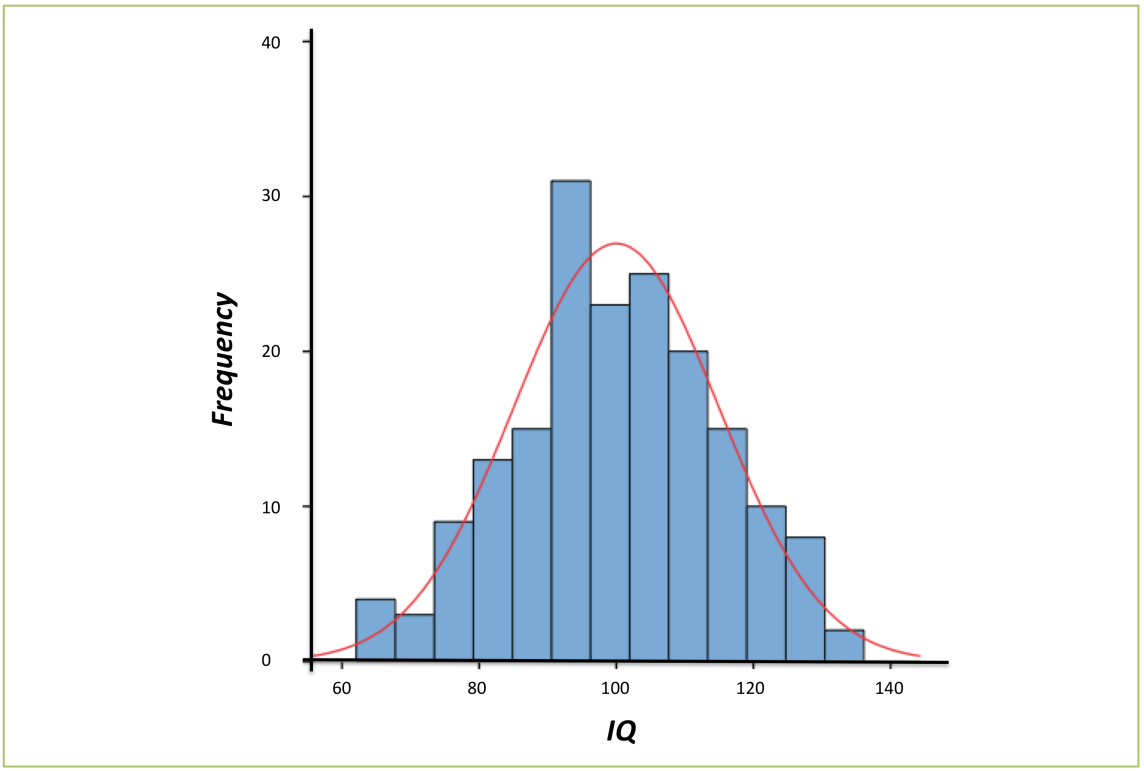


Figure 5.7. Distribution of IQ scores for all participants (N = 169).

Finally, 56 females were selected and divided in two groups: training (N = 28) and control (N = 28). Members of each group were carefully matched for their general intelligence index (see Figure 5.8).

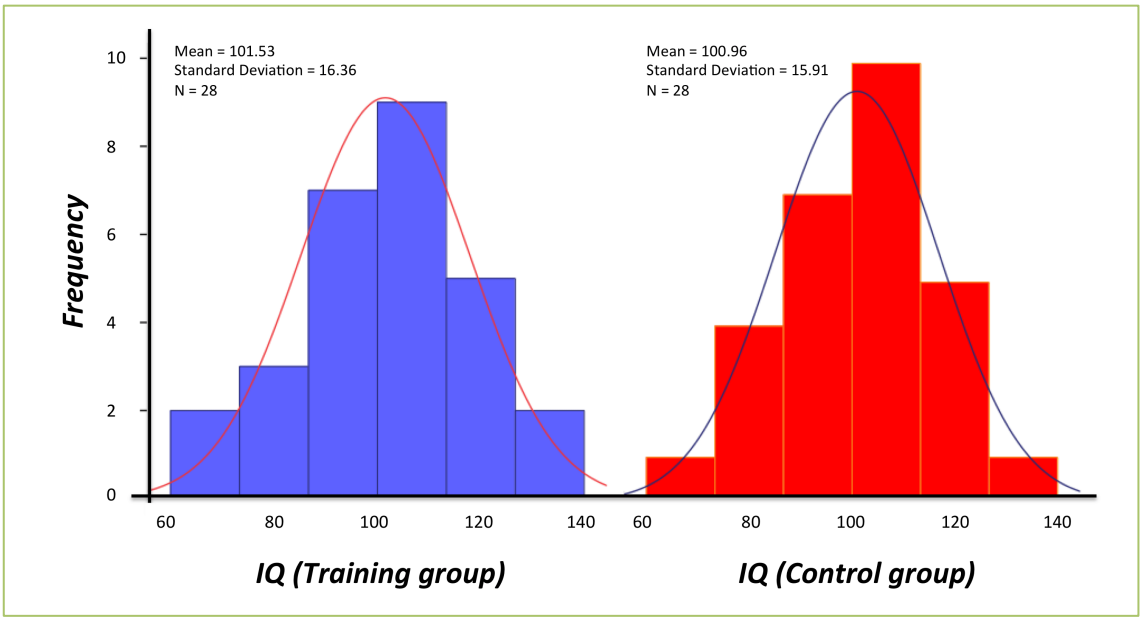


Figure 5.8. Distribution of IQ scores for control (red) and training (blue) groups.

Participants were paid for their participation. Each participant in the training group received 200€, while each control subject received 100€. They signed an informed consent (Appendix 3), following the Helsinki guidelines (World Medical Association, 2008). The local ethics committee approved the study.

5. 4. Stage 3: First scan.

MRI details were described in Chapter 2. Now we provide the specific parameters employed in this study.

5. 4. 1. *Structural images acquisition.*

Structural images (T1-weighted) were acquired in a General Electric Signa 3 Tesla MR Scanner (General Electric Healthcare, Farfield, CT) using a whole-body radiofrequency (RF) coil for signal excitation and quadrature 8-channel coil for reception. For the structural images analyzed here, a high-resolution 3D T1-weighted Gradient Echo-SPGR was applied, with parameters: TE = 4.1 ms, TR = 9.1 ms, flip angle = 10°, 170 sagittal slices, acquisition matrix = 256 mm x 256 mm, isotropic voxel size = 1 mm^3.

5. 4. 2. *Diffusion-weighted images acquisition.*

Diffusion-weighted images were also acquired for each participant. The pulse sequence was single-shot, diffusion-weighted, echo planar acquisition with 45 diffusion-sensitized gradient directions and 2 b0 images (raw T2 signal with no diffusion weighting) with the following parameters: TR =12700 ms, TE = 84.4, flip angle = 90, NEX = 2, 50 oblique AC-PC (Anterior-Posterior commissure) oriented slices, acquisition matrix = 96 mm x 96 mm, slice thickness = 2.4 mm, in-plane resolution = 0.94 mm2; b value = 1000 s/mm^2.

Figure 5.9 shows examples of the MRI images acquired: T1-weighted, b0 and diffusion-weighted images.

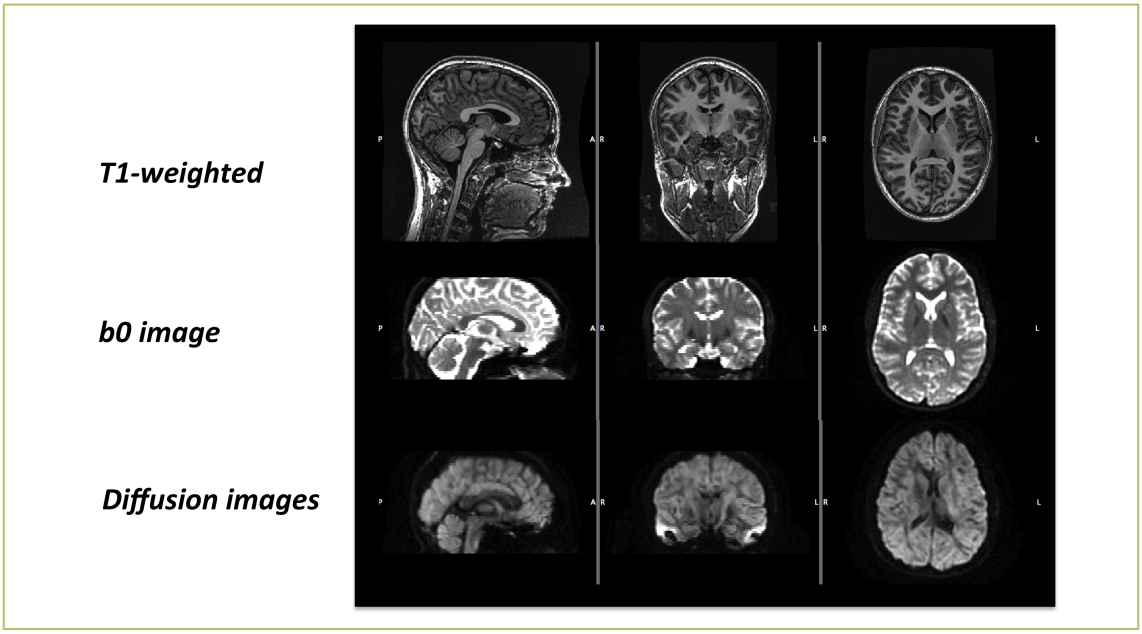


Figure 5.9. Examples of T1-weighted, b0 and diffusion-weighted (45 gradient directions) images.

5. 5. Stage 4: Adaptive n-back training program.

The training group (N = 28) completed the cognitive program based on the dual n-back task (Jaeggi et al., 2008) on individual cabins under strict supervision. The training program lasted for twelve weeks and twenty-four sessions (approx. 30 min per session). In the first 8 sessions, single visual and auditory versions of n-back task were used. The first 4 sessions were devoted to the visual modality, and the next 4 sessions to the auditory modality. Afterwards, participants completed sixteen dual sessions (visual +

auditory). Data were analyzed every week to check the progress at both the individual and the group level. Participants received systematic feedback regarding their performance.

The control group was passive. Members of this no-contact control group were invited to follow their normal life as university students since the present research was not aimed at comparing different types of training. Note that previous studies (Chooi & Thompson, 2012; Redick et al., 2013; Karbach & Verhaeghen, 2009) failed to find differences between active and passive control groups. Furthermore, published meta-analyses did not detect any difference between placebo and no-contact groups (Klauer & Phye, 2008; Au et al., 2014).

5. 5. 1. *Visual n-back (single task).*

Each session included 12 blocks and each block comprised 20 trials + n trials depending on the n-back level (e.g. difficulty n-back level 4 = 24 trials). Blue squares were pseudo-randomly displayed in 8 different positions on the screen. Each square was presented during 500 ms and the participant could provide the answer once the square appears and during 2500 ms. Therefore, the inter-stimulus interval was 3000 ms (See Figure 5.10a).

Targets were shown 6 times within each block. Participants had to press the "Z" key when the target was presented in the same position than "n" trials before. Hits, errors of commission (false alarms) and errors of omission (don't press the bottom when target is showed) were computed by block. The first block of the first session began at difficulty level 1 ("press key "Z" when the square was presented in the same position than one before"). From there, the level of difficulty was adapted according to participants' performance on each block. The rules for increasing, decreasing or keeping the level of difficulty in the next block were:

- Increasing: less than 3 errors (commission + omissions) in the block.
- Decreasing: more than 5 errors (commission + omissions) in the block.
- Keep the level of difficulty: between 3 and 5 errors (commission + omissions) in the block.

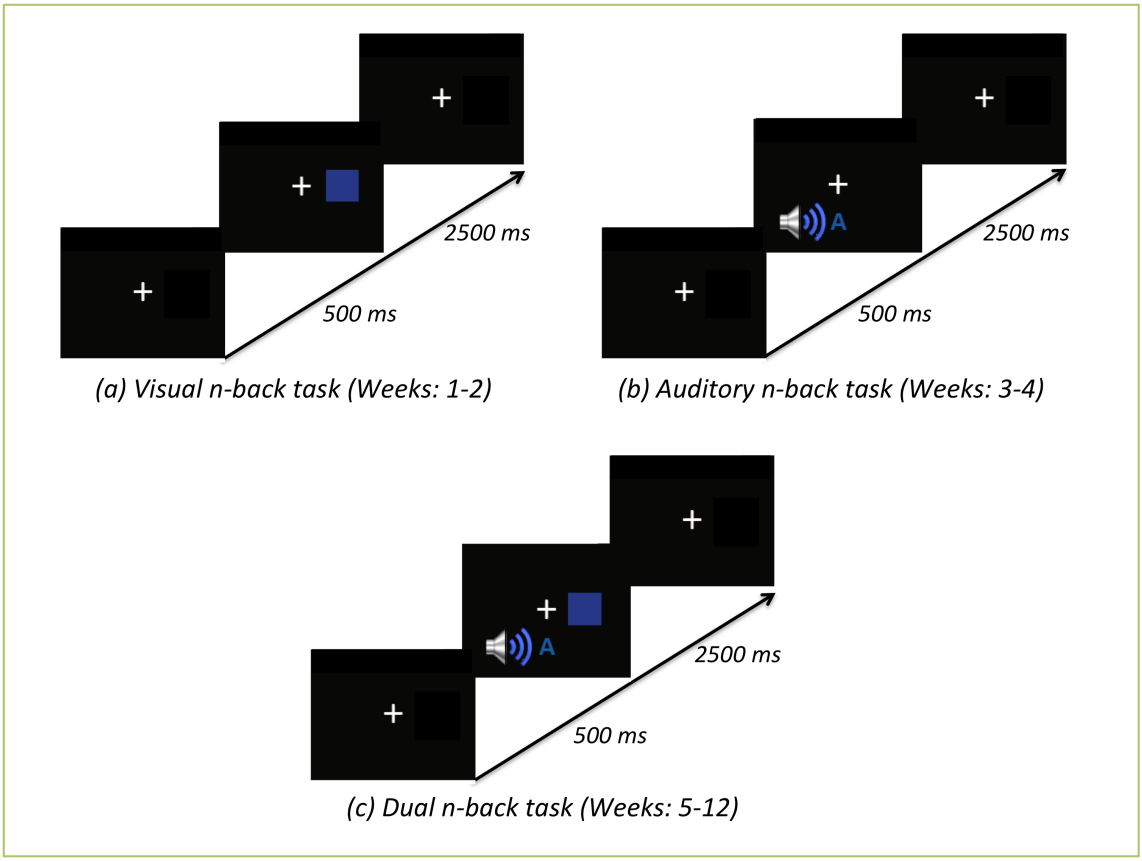


Figure 5.10. N-back training program: (a) Visual n-back, (b) Auditory n-back and (c) Dual n-back.

The level of difficulty obtained in block 12 (the last block of each session), was set as the level of difficulty of the first block in the next session. Finally, at the end of each session, and at the beginning of a new session, the participant received feedback regarding her performance.

5. 5. 2. *Auditory n-back (single task).*

Each session included 12 blocks and each block comprised 20 + n trials (where n depended on the n back level following the same rule than for the visual version). The stimuli were the letters A, E, G, I, K, O, T, and Q. The participant listened to the sequence of letters via headphones. The sequence (See Figure 5.10.b) and rules for increasing or decreasing the level were the same than for the visual version. Now participants must press "M" when they listen to the same letter than "n" trials before.

5. 5. 3. *Dual n-back.*

Each session included 12 blocks with 20 + n back trials per block, where visual and auditory stimuli were presented simultaneously (See Figure 5.10.c). The n-back level for both tasks (visual and auditory) was the same for each level. The level was not allowed to change for one task only. Stimuli were the same than for the single tasks (visual and auditory). The participant had to press the "Z" key for visual stimuli and the "M" key for auditory stimuli when the square displayed on the screen or the letter they heard were the same than "n" trials before. Again, targets for both conditions were presented pseudo-randomly. It could or could not happen that targets for visual and auditory conditions showed-up in the same trial (See Figure 5.11).

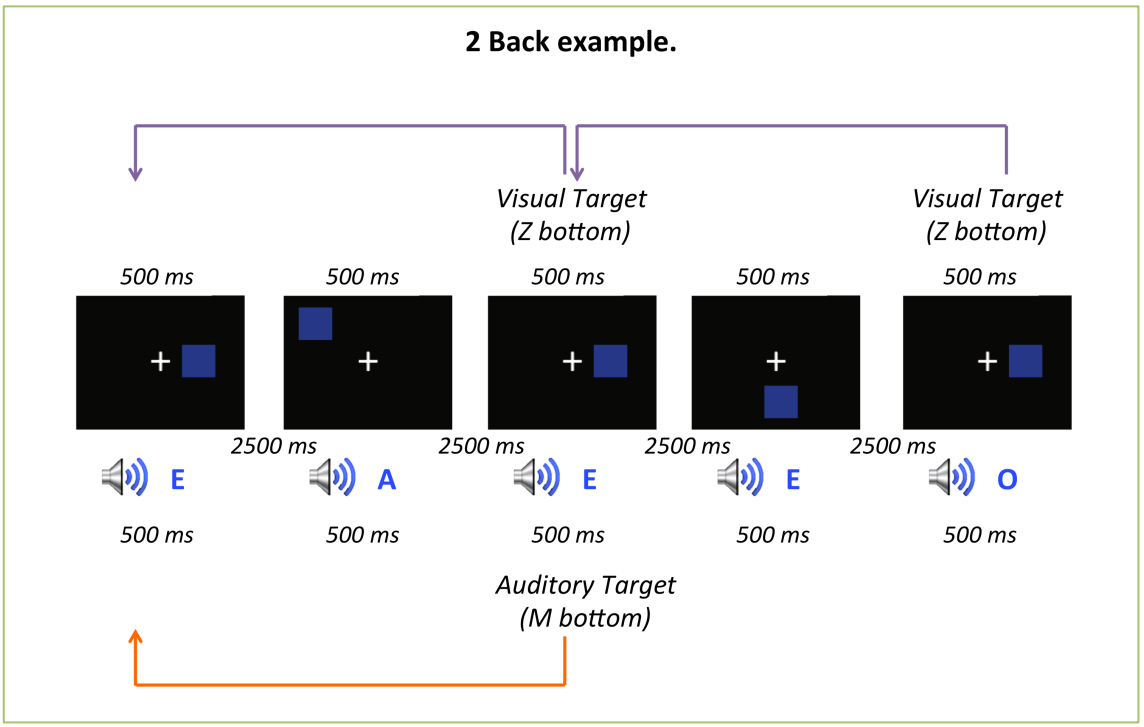


Figure 5.11. Example 2-back level in dual n-back task. The letters were auditory presented at the same rate as the spatial material that was presented visually. Participants had to press the "Z" button for visual targets and "M" button for auditory targets.

Targets were presented 6 times per block. The rules for increasing or decreasing the level after a block were the same than for the single n-back task. However, errors for both conditions (visual + auditory) were considered simultaneously.

A motivation questionnaire (Tapia & de la Red Fadrique, 2007) was completed every two weeks. They were asked about their (a) involvement in the task, (b) perceived

difficulty level, (c) perceived challenge of the task levels, and (d) expectations of future achievement. At the end of the training period, participants were asked about their general evaluation of the program, using a rating scale from very low (0) to very high (10). To maintain our participants' motivation, a graph comparing their performance with other teams (hypothetical sample of American students) was shown at the midpoint of the dual n-back training program.

5. 6. Stage 5: Second scan.

All participants (N = 56) were scanned again after completing the training, using the same MRI protocol (stage 3).

5. 6. 1. *Image Preprocessing (T1-weighted images)*

Firstly, images were converted from the DICOM format (Digital Imaging and Communication in Medicine) to NIFTI (Neuroimaging Informatics Technology Initiative). Secondly, scan pairs for the same participant were carefully inspected. A warping distortion between the pre-test and post-test scans was observed, namely, a stretch expansion in the temporal lobe area and the opposite in the parietal area. This distortion was due to (a) the use of high field strength in the scanner and (b) the offset of iso-center in the longitudinal scans. Therefore, recommended corrections were applied using a tool called "Grinder". This tool corrects for geometric distortions due to uneven field strength within the scanner. The applied image pre-processing steps for this correction involved Grinder + N3 + bias correction from SPM5 (Statistical Parametric Mapping 5) unified segmentation, which is similar to the protocol applied in ADNI (Jack et al., 2008). These steps were successful with the exception of two participants from each group. Following the advice of Paul M. Thompson, this quality control was completed in Mayo Clinic (USA) and supervised by Jeff Gunter.

5. 6. 2. *Image Preprocessing (Diffusion-weighted images)*

As in T1 preprocessing, diffusion weighted images (DWI) were transformed from DICOM format to NIFTI format. DWI images were pre-processed using the FMRIB's Diffusion Toolbox (FDT). Correction for motion and geometrical distortion due to Eddy currents was performed with the *eddycorrect* tool in FDT, taking as the reference image

the average of the two b0 volumes. Non-brain tissue from the average b0 image was removed using the FMRIB's Brain Extraction Toolbox, BET (Smith, 2002).

5. 7. Stage 6: Second psychological assessment.

The second psychological assessment was completed in individual sessions for all participants (N = 56). The same intelligence tests were administered, but now the even numbered items were presented (odd items were administered in the first psychological assessment). The same cognitive tasks were administered in individual cabins.

5. 8. Stage 7. Follow-up: Adaptive n-back (faces).

The follow-up was focused on near-transfer and far-transfer. Participants (N = 56) were contacted for completing a session one-year after ending the training program (1 April – 13 April, 2013). 22 women from the training group and 15 from the control group accepted the invitation (Total = 37). This follow-up was intended to study performance of the control and training groups on a parallel n-back task (near transfer). The assessment was strictly equivalent. The near-transfer task was based on stimuli different to those considered in the original training program (Román et al., 2015) (see Figure 5.12).

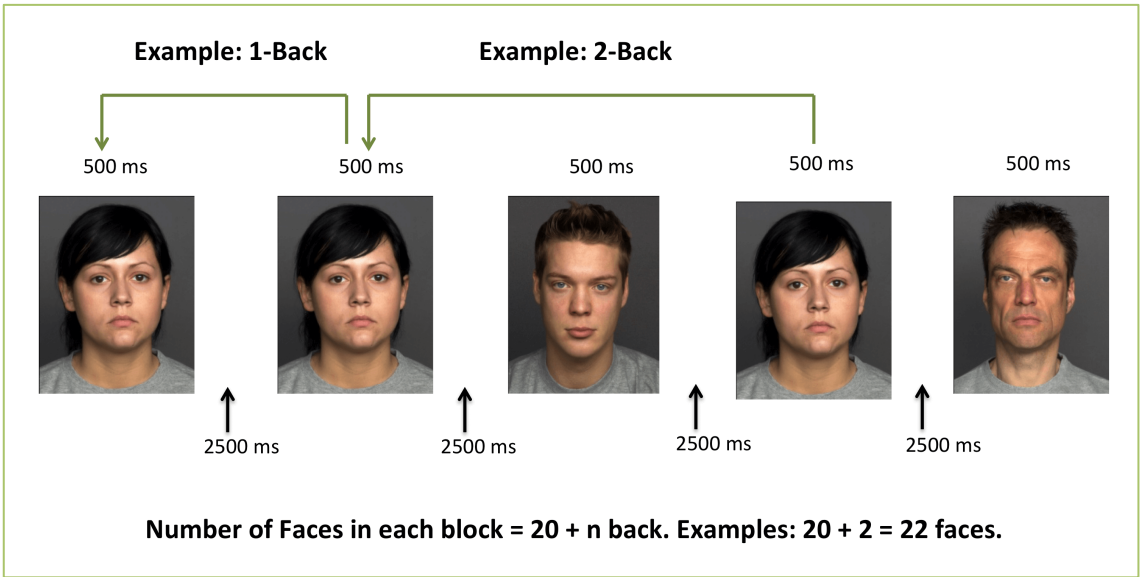


Figure 5.12. Example of faces (n-back) task (follow-up).

The follow-up task was similar to the visual n-back task. However, while the visual n-back task required spatial processing, faces n-back task did not. In the latter instance, only updating processes are required.

5. 9. Stage 8. Follow-up: Academic grades.

Far-transfer were analyzed using academic grades on several courses at different time points with respect to the training program: (a) Statistics (3 months after the training program), (b) Psychometrics (2 years after the training program), and (c) Psychological Treatments (3 years after the training program). We chose these course topics because the considered participants completed the same standardized academic knowledge test.

CHAPTER 6: Training performance (Q1).

6.1. Introduction.

This chapter includes four sections:

(a) *N-back performance and relationships with the psychological factors*. We expected improvement during the three modalities of cognitive training program in the n-back level achieved, similarly to previous works (e.g. Chooi & Thompson, 2012; Jaeggi et al., 2008; 2014; Redick et al., 2012). Also, we hypothesize that those psychological factors requiring reliable preservation of relevant information in the short-term (fluid intelligence –Gf– and working memory capacity –WMC–) will predict the n-back performance across training sessions (Jaeggi et al., 2010a,b) because Gf and WMC share the short-term component (Martínez et al., 2011). Personality traits will be less correlated than cognitive factors, although neuroticism and conscientiousness may also be relevant (Studer-Luethi et al., 2012).

(b) *Transfer effects associated with the training program*. Regarding this issue, our hypothesis is that transfer will be more visible in factors directly related with n-back (WMC; *near-transfer*) and probably will be extended to fluid intelligence (Gf, *far-transfer*), because WMC and Gf share cognitive processes (Colom et al., 2004, 2005, 2008; Jaeggi et al., 2008, 2010a,b, 2011; Halford et al., 2007; Kyllonen & Christal, 1990; Martínez et al., 2011).

(c) *Influence of achieved performance on the training program* (standard and top performers). Probably, top performers will show higher intelligence scores from the outset, as n-back tasks can be used to measure individual differences in Gf (Jaeggi et al., 2010b). Also, we will check whether the performance in training is related to the observed transfer effects (see Jaeggi et al., 2011).

(d) The last section is dedicated to discussing the findings. The analyses reported here were circumscribed to participants passing quality control (QC) of imaging processing of T1-weighted images (N = 26) (see Stage 5 in chapter 5). Analyses with the full group can be found in Colom et al. (2013a; see Appendix 4).

6.2. N-back performance.

First, improvements across training sessions were analyzed considering the average n-back level achieved in visual, auditory, and dual training sessions (Figure 6.1). Repeated measures analysis of variance (ANOVA) was computed for each modality of training. The dependent variable was the average n-back level computed in each session (4 levels in visual and auditory modalities and 12 in dual n-back). The Greenhouse-Geisser correction was applied when the sphericity assumption was violated. Partial eta squared (η_p^2) was computed as measure of effect size.

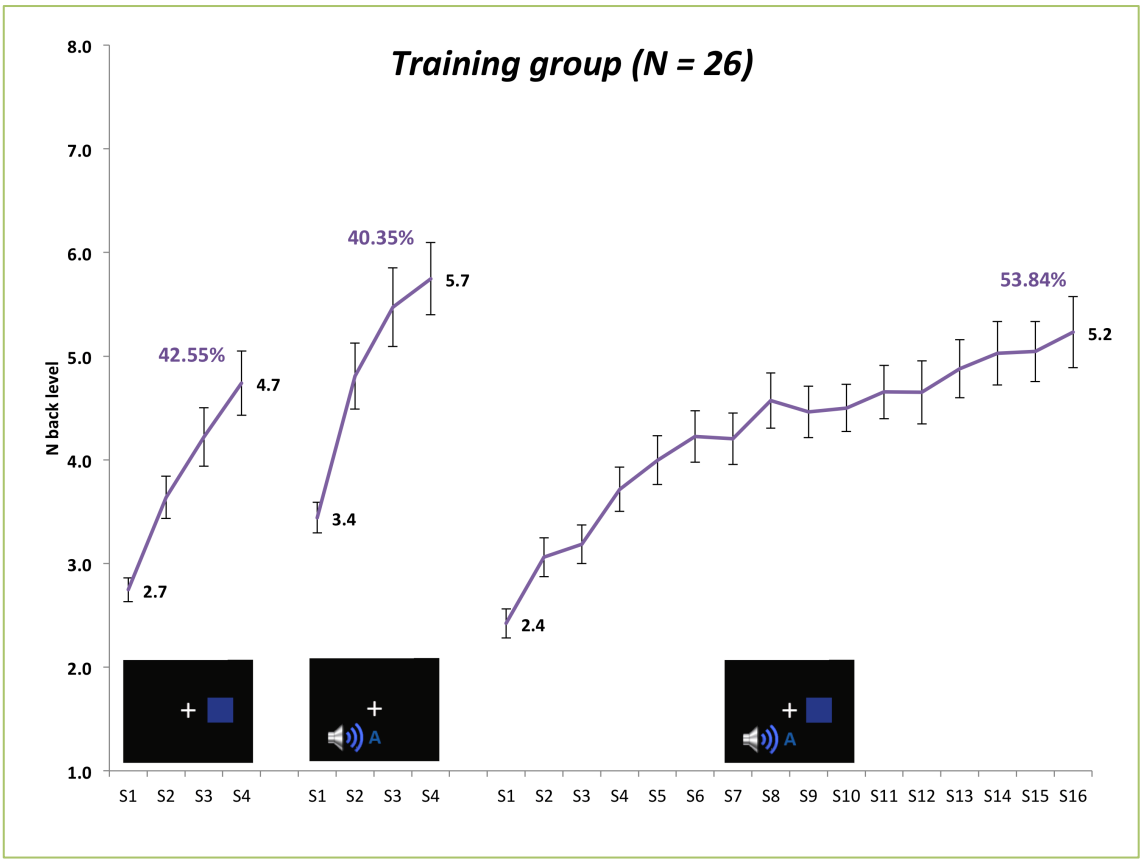


Figure 6.1. Performance on the n-back training program (N = 26). From left to right: visual, auditory and dual n-back task. S = Session.

A significant main effect of session was found for all conditions: $F(3,75) = 33.114$, $p < .001$, $\eta_p^2 = .570$ (visual), $F(3,75) = 40.687$, $p < .001$, $\eta_p^2 = .619$ (auditory) and $F(15,375) = 40.345$, $p < .001$, $\eta_p^2 = .617$ (dual). Therefore, on average, participants improved their performance across sessions.

Next, the percentage of improvement was computed according to Chooi and Thompson (2012).

$$\%\ \boldsymbol{Improvement} = \frac{\boldsymbol{Avg.Highest\ Training\ score - Avg.First\ Training\ score}}{\boldsymbol{Avg.Highest\ score}}\ \boldsymbol{x\ 100} \qquad [1]$$

The percentage of improvement was highest for the dual task (53.84%) followed by the visual (42.55%) and auditory tasks (40.35%).

The correlation between n-back performance across sessions and the cognitive factors measured in the first psychological assessment (before training) was calculated (Figure 6.2).

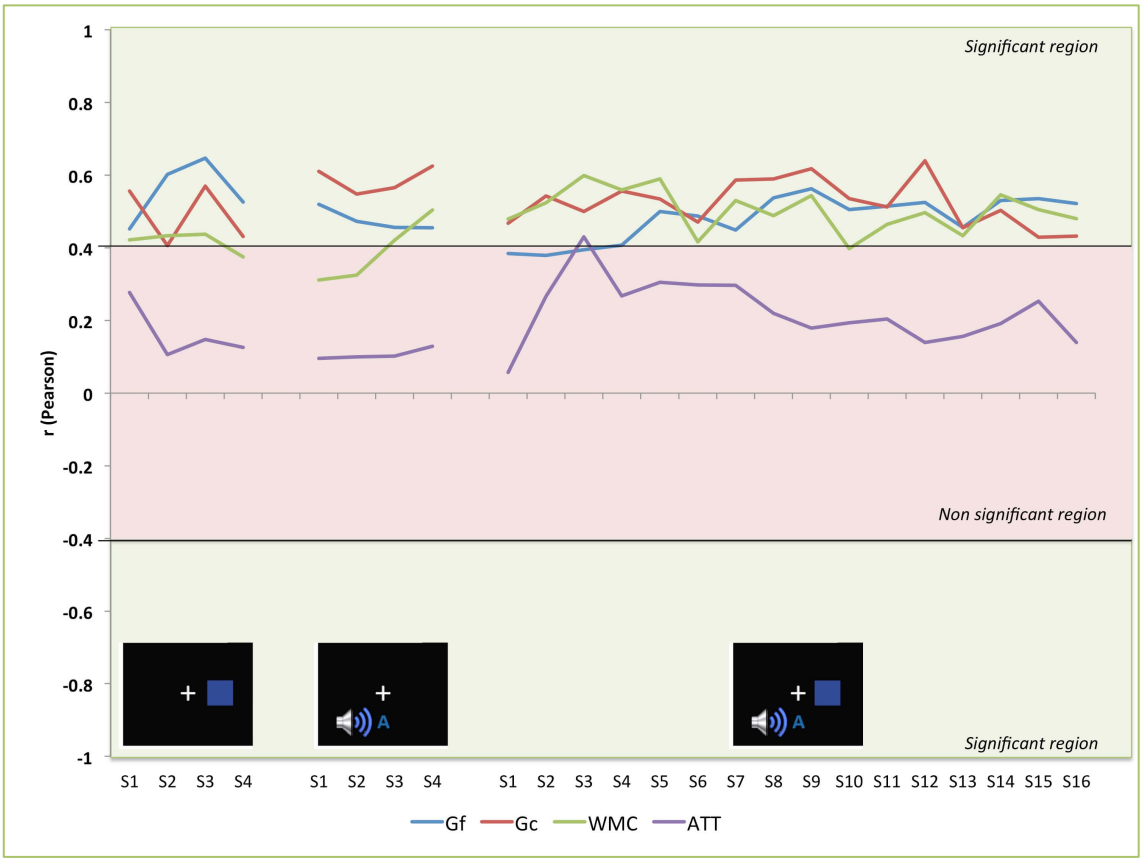


Figure 6.2. Correlation between intelligence/cognitive factors before training and achieved n-back levels across training sessions. Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity and ATT = attention control. S = session. From left to right: visual, auditory and dual n-back tasks.

Correlations with fluid intelligence (Gf), crystallized intelligence (Gc) and working memory capacity (WMC) were statistically significant (higher than .40). Interestingly,

the correlation with Gf showed an increased tendency across sessions. Attention showed a lack of correlation with training performance.

Correlations between personality traits and training performance are shown in Figure 6.3.

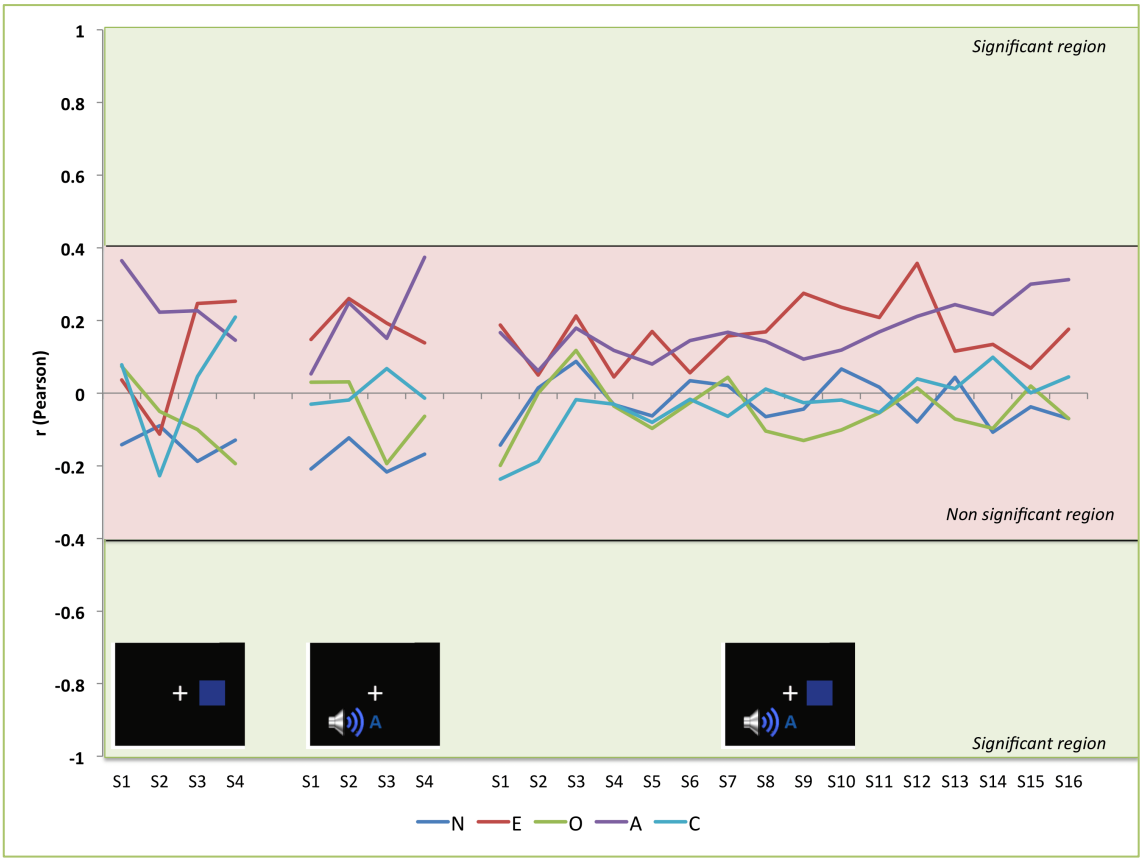


Figure 6.3. Correlation between personality traits and achieved n-back levels across training sessions. N = Neuroticism, E = Extraversion, O = Openness to experience, A = Agreeableness, C = Conscientiousness and S = Session. From left to right: visual, auditory and dual n-back tasks.

Individual differences in key personality traits were not related with individual differences in n-back performance.

Regarding motivation (possible range of scores: 0-10), the average values computed from the completed questionnaire in (visual and auditory) session 4, as well as in (dual) sessions 4, 8, 12 and 16 were 8.08 for 'involvement with the task' (range 7.88-8.23), 7.86 for 'perceived difficulty level' (range 7.35-8.35), 8.08 for 'perceived challenge of task levels' (range 7.81-8.35) and 6.92 for 'expectations for future achievement' (range 6.54-7.27). The mean for the general evaluation was 7.73.

6.3. Transfer effects.

Supplementary Tables 2a and 2b show scores before and after n-back training program for the training and control groups, respectively.

As noted before, odd items in the standardized intelligence tests were administered before n-back training while even items were administered after training. For equating the difficulty levels of odd and even items, item response theory (IRT) was applied (details can be found Appendix 4, A.3. section). Briefly, three steps were followed: (a) independent samples were considered for calibrating the items, obtaining IRT odd and even item parameters using the same metric, (b) with item parameters fixed to those obtained in the previous stage, IRTPRO software was applied separately to the odd test and the even test, obtaining two conversion tables. Each conversion table indicated what level of ability (θ) corresponded to each sum score in this part. The same prior distribution for θ (mean = 0; standard deviation = 1) was assumed when computing the conversion table; finally (c) conversion tables were applied to transform the sum scores in IRT scores for the training and control groups. The obtained IRT scores were considered for analyzing transfer effects with respect to intelligence.

Figure 6.4 shows the scores for each cognitive factor (Gf, Gc, WMC and ATT) before and after training in the two groups of participants.

Afterwards, standardized change scores (Jaeggi et al., 2011) were computed at both the construct and measure level. The standardized change was computed using the folllowing formula:

Stand. Change = (score after – score before)/ SD of scores before. [2]

These scores were submitted to analysis of covariance (ANCOVA), where the dependent variable was the standardized score, the independent variable was the group, and the covariate was the score before training. A *p* level of .05 (one-tailed) was considered (Jaeggi et al., 2008). Results at the construct level are shown Figure 6.5 (note that ATT scores were reversed for clarity).

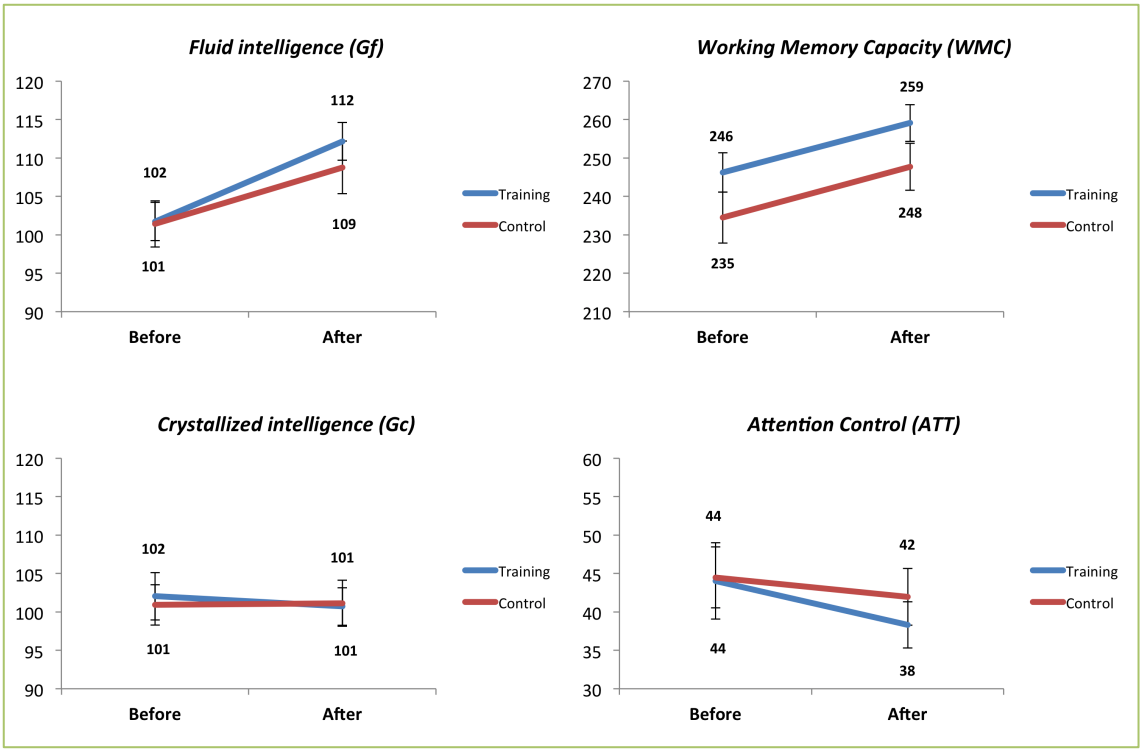


Figure 6.4. Cognitive scores before and after n-back training program in the training (N = 26) and control (N = 26) groups at the construct level.

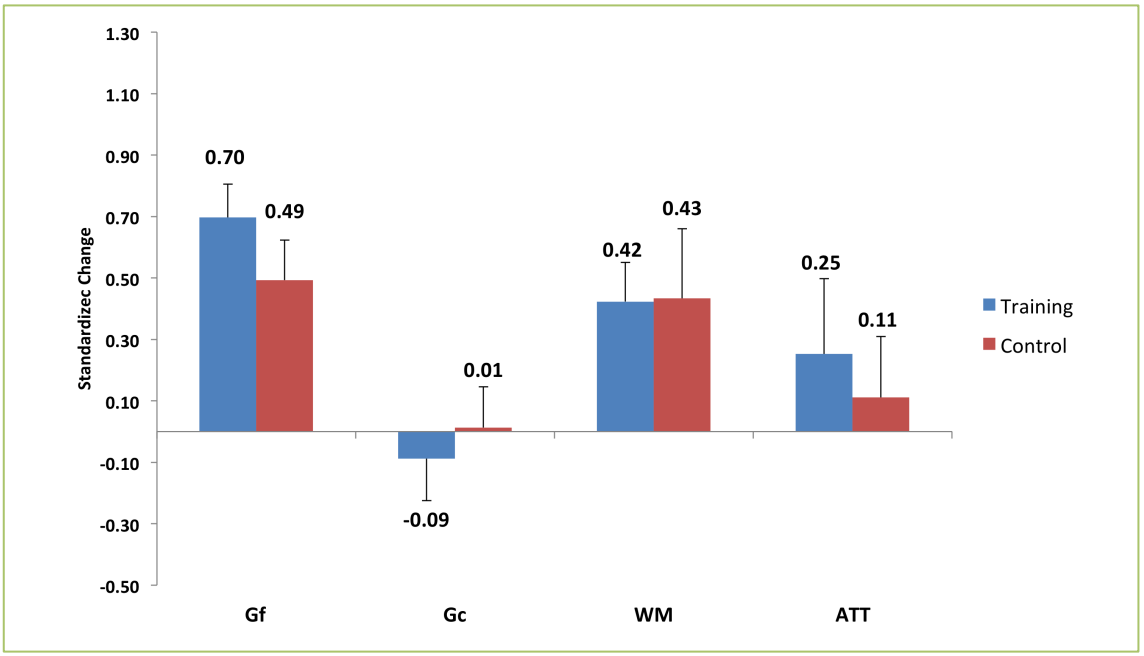


Figure 6.5. Standardized change [(after − before)/ SD before] at the construct level in the training and control groups. Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity, ATT = attention control.

No statistically significant group differences were observed before training. Gf and WMC scores increased in both groups in the posttest, ATT scores showed a decrease (less conflict), and Gc remained at the same level.

These results indicate (a) increased Gf, WMC, and ATT scores in both groups, and (b) null changes for Gc in both groups. No statistically significant group differences were found, $F(1,49) = 1.507$, $p = .113$, $\eta_p^2 = .030$ (Gf); $F(1,49) = .083$, $p = .325$, $\eta_p^2 = .004$ (Gc); $F(1,49) = .708$, $p = .202$, $\eta_p^2 = .014$ (WMC) and $F(1,49) = .579$, $p = .225$, $\eta_p^2 = .012$ (ATT). Therefore, we failed to find transfer effects at the construct level.

Afterwards, the results at the measure level were analyzed. Figure 6.6 shows these results (left panel for the intelligence tests and right panel for the cognitive tasks).

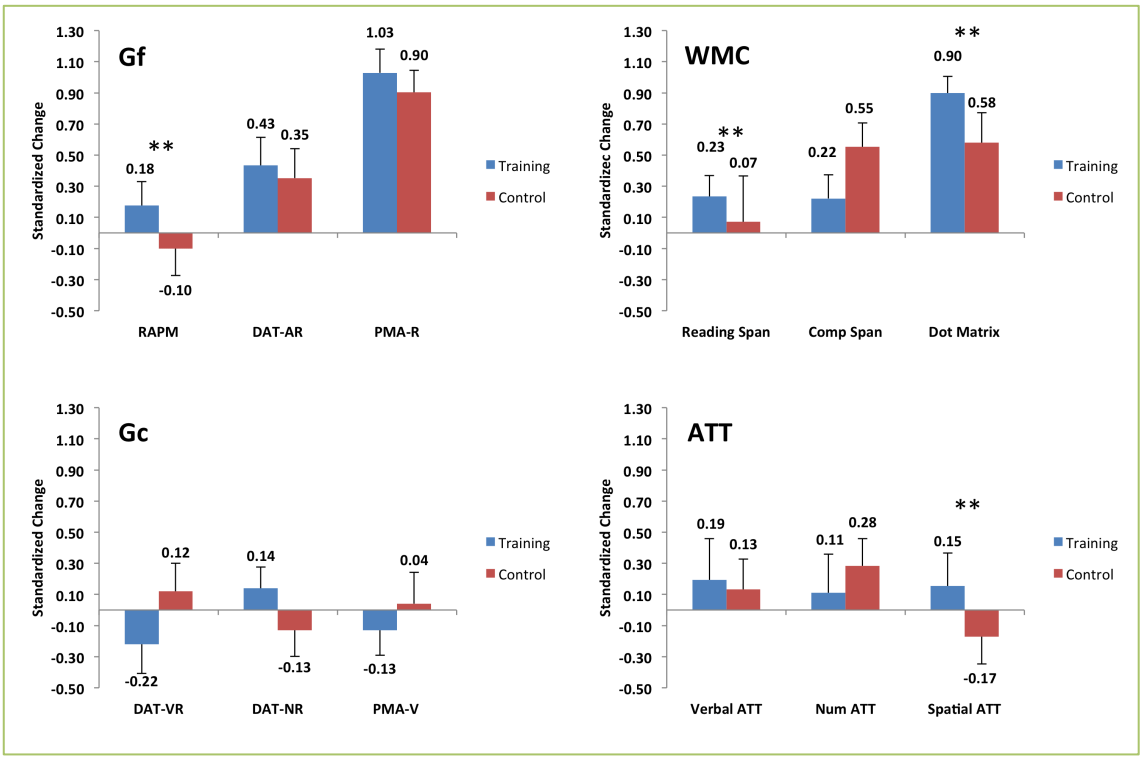


Figure 6.6. Standardized change [(after − before)/ SD before] in the training and control groups at the measures level: Gf = fluid intelligence (RAPM = Raven Advanced Progressive Matrices Test, DAT-AR = abstract reasoning, PMA-R = inductive reasoning); Gc = crystallized intelligence (DAT-VR = verbal reasoning, DAT-NR = numerical reasoning, PMA-V = vocabulary); WMC = working memory capacity; ATT = attention control (verbal ATT = vowel–consonant, numerical ATT = odd–even, spatial ATT = right–left). .** = Significant difference between groups.

The group difference in RAPM favoring the training group was close to statistical significance, $F(1,49) = 2.217$, $p = .071$, $\eta_p^2 = .043$. Results for DAT-AR were not

statistically significant, $F(1,49) = .010$, $p = .460$, $\eta_p^2 = .001$. There was a large change for both groups in the highly speeded Gf test (PMA-R), amounting to one standard deviation in training group and 0.90 in control group, $F(1,49) = .611$, $p = .219$, $\eta_p^2 = .012$.

Changes observed in Gc measures were not statistically significant: $F(1,49) = .996$, $p = .161$, $\eta_p^2 = .020$ (DAT-VR), $F(1,49) = .583$, $p = .224$, $\eta_p^2 = .012$ (DAT-NR) and $F(1,49) = .015$, $p = .451$, $\eta_p^2 = .001$ (PMA-V).

The training group showed statistically significant increases in reading span, $F(1,49) = 4.907$, $p = .015$, $\eta_p^2 = .091$ and dot-matrix, $F(1,49) = 2.412$, $p = .047$, $\eta_p^2 = .051$. The difference in computation span was no statistically significant, $F(1,49) = 1.766$, $p = .095$, $\eta_p^2 = .035$.

Finally, the training group showed statistically significant improvements in spatial attention (Simon task), $F(1,49) = 3.946$, $p = .026$, $\eta_p^2 = .075$. Changes in numerical and verbal attention were not statistically significant: $F(1,49) = .002$, $p = .482$, $\eta_p^2 = .001$ (Numerical Flankers), $F(1,49) = .550$, $p = .231$, $\eta_p^2 = .011$ (Verbal Flankers).

6.4. Top and standard achievers.

There are substantial individual differences in training performance, which may have had some impact on transfer effects (Jaeggi et al., 2011, 2014). Hence, the training group was divided according to the mean obtained in last training session of the dual n-back task: (a) standard performance (n-back < 5.23; N = 16), and (b) top performance (n-back > 5.23; N = 10). This approach, employed in previous reports (Jaeggi et al., 2011), has the benefit of rendering the results readily interpretable, allowing comparisons between top and standard performers. However, this method loses some power by categorizing a continuous variable.

The data were re-analyzed in order to answer these questions:

- Can these groups be distinguished in the cognitive constructs measured in the first psychological assessment (before training)?

- When (which session) did the differences start between both groups in the achieved n-back?

- *Are transfer effects sensitive to training performance?*

6.4.1. *Individual differences before training.*

The hypothesis was that standard and top performers should be different in the psychological factors related with n-back performance (Gf, Gc and WMC). This should not be the case for ATT and personality traits. Therefore, independent sample T-tests were computed for each psychological factor (cognitive and personality traits). Also, effect sizes (Cohen *d*; Cohen, 1992) were computed.

Figure 6.7 shows cognitive results whereas Figure 6.8 depicts results for the personality traits.

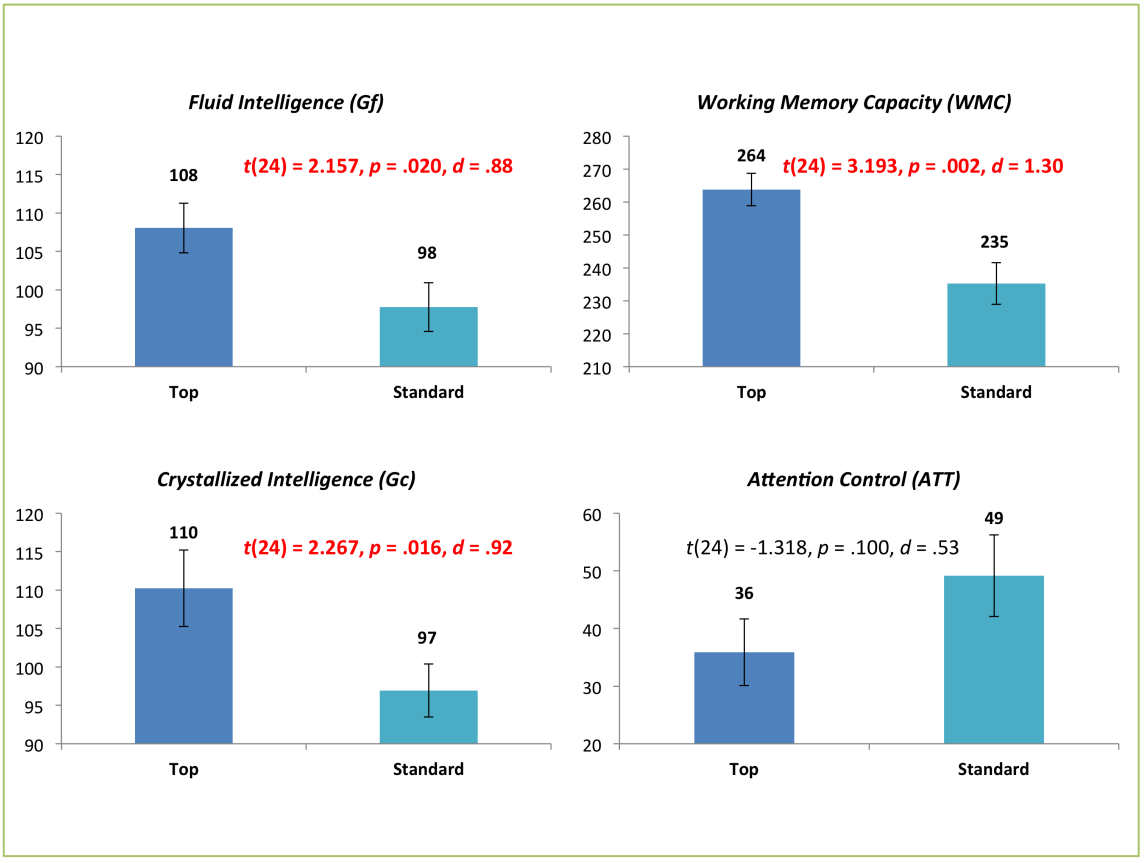


Figure 6.7. Cognitive differences between top training (N = 10) and standard training (N = 16) before n-back training program.

As hypothesized, there were large group differences (*d* Cohen > .80) between top and standard training groups in the cognitive factors related with training performance. The highest difference was found for WMC (*d* = 1.30) followed by Gc (*d* = .92) and Gf (*d* = .88). Although the difference for ATT was not significant, the effect size suggests a

moderate effect (*d* Cohen = .53). Again, as hypothesized, we did not find statistically significant differences between groups with respect to the measured personality traits [N: $t(24) = -.39$, $p = .699$, $d = .16$; E: $t(24) = .94$, $p = .358$, $d = .38$; O: $t(24) = -.20$, $p = .845$, $d = .08$; A: $t(24) = .44$, $p = .666$, $d = .18$; C: $t(24) = -.04$, $p = .970$, $d = .01$].

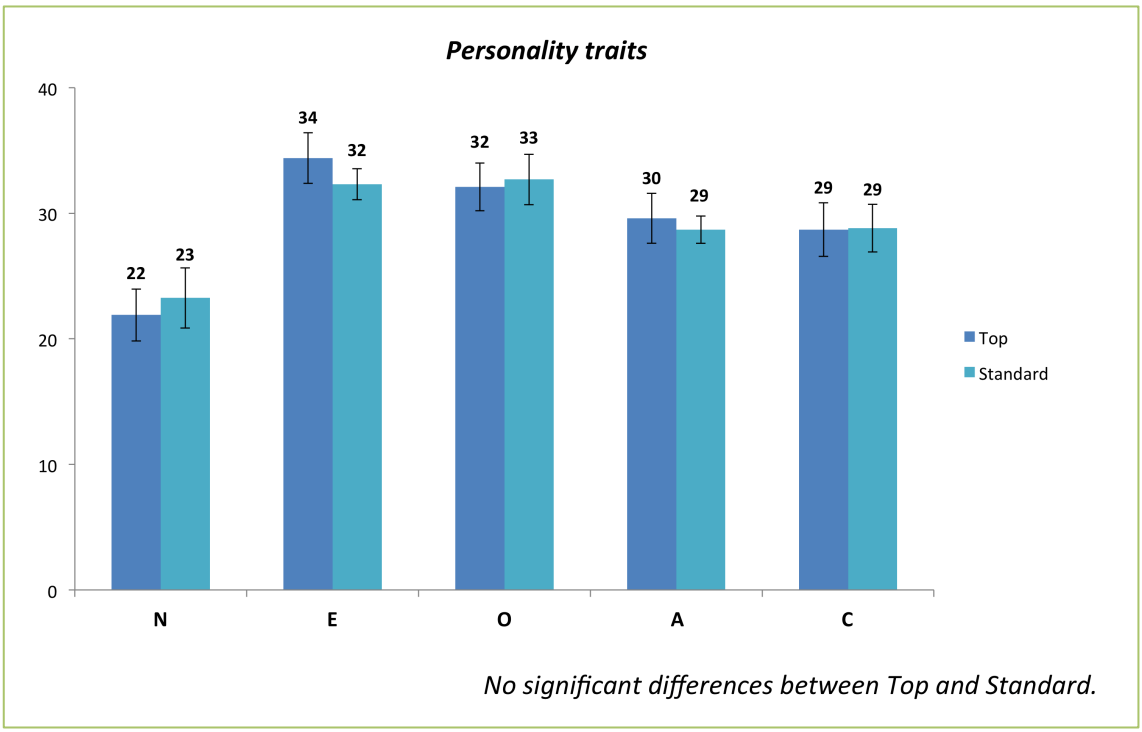


Figure 6.8. Personality differences between top training (N = 10) and standard training (N = 16) before the training program. N = neuroticism, E = extraversion, O = openness to experience, A = agreeableness, C = conscientiousness.

6.4.2. *N-back performance differences.*

Figure 6.9 shows average levels across training sessions in top and standard performers.

A 2x4 ANOVA (Group x Sessions) was computed to verify whether these groups were different in their performance on the visual and auditory n-back tasks. A 2x16 ANOVA (Group x Sessions) was computed for the dual n-back task. Group was a between factor and Session was a repeated measures factor. The Greenhouse-Geisser correction was applied when the sphericity assumption was violated. Partial eta squared (η_p^2) was computed as measure of effect size and Bonferroni correction was applied. Significance level was set at $p < .05$.

Main effects of Sessions [$F(3,72) = 39.736$, $p < .001$, $\eta_p^2 = .623$] and Group, [$F(1,24) = 10.843$, $p = .003$, $\eta_p^2 = .311$] were found in the visual n-back task. The effect of their interaction was also significant [$F(3,72) = 3.499$, $p = .020$, $\eta_p^2 = .127$]. Post-hoc analyses showed that both groups were different in all sessions ($p < .05$). Results for the auditory n-back task were similar (main effects of Sessions [$F(3,72) = 58.584$, $p < .001$, $\eta_p^2 = .709$], main effect of Group, [$F(1,24) = 17.757$, $p < .001$, $\eta_p^2 = .425$] and effect of interaction [$F(3,72) = 8.094$, $p < .001$, $\eta_p^2 = .252$]).

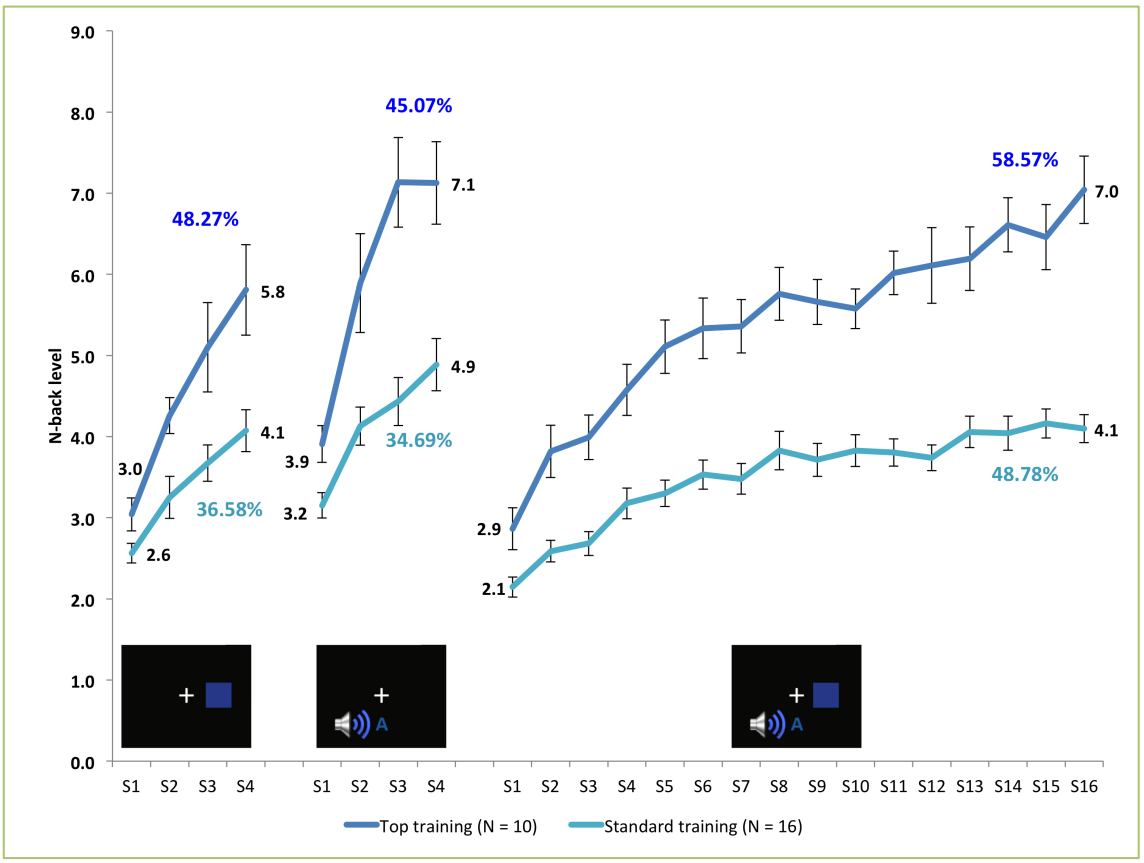


Figure 6.9. Performance for the standard and top groups. Dark blue line shows performance in n-back for top training participants (N = 10) and blue line depicts performance in n-back for standard training participants (N = 16).

Post-hoc analysis showed statistically significant group differences in all sessions ($p < .05$). Finally, results remained the same for the dual n-back task (main effects of Sessions [$F(15,360) = 52.188$, $p < .001$, $\eta_p^2 = .685$], main effect of Group, [$F(1,24) = 50.916$, $p < .001$, $\eta_p^2 = .680$] and effect of interaction [$F(15,360) = 5.477$, $p < .001$, $\eta_p^2 = .186$]). The group difference was significant in all sessions ($p < .05$, Bonferroni comparisons).

The percentage of improvement was also computed according to Chooi and Thompson's formula (Chooi & Thompson, 2012). The results are shown in Figure 6.9. The top group showed an increment statistically greater than the standard group and the difference was 10% (or higher) for all tasks.

Finally, the mean number of errors for all blocks in each session (omission + commission) was computed. The Greenhouse-Geisser correction was applied when the sphericity assumption was violated. Partial eta squared (η_p^2) was computed as measure of effect size and Bonferroni correction was applied. Significance level was set at $p < .05$. Figure 6.10 shows the mean number of errors across blocks in the single n-back tasks.

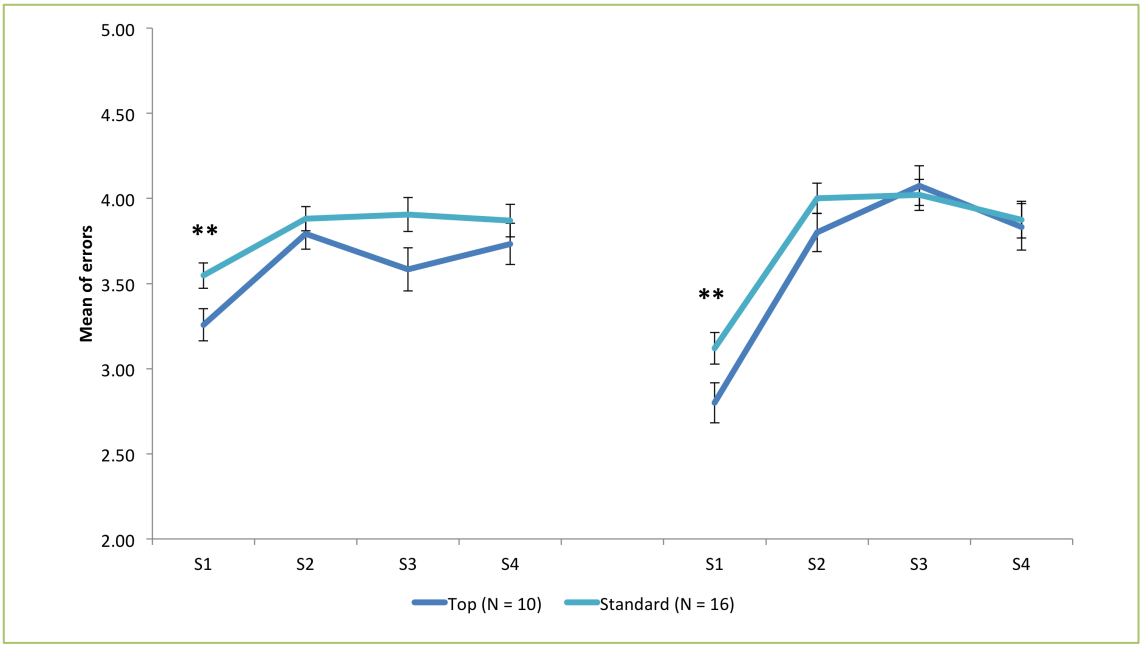


Figure 6.10. Dark blue line depicts mean of errors for top training participants (N = 10) and blue line depicts mean of errors for standard training participants (N = 16). From left to right: visual and auditory tasks.** = Significant difference between groups.

A main effect of Session [$F(3,72) = 7.702$, $p < .001$, $\eta_p^2 = .243$] and Group [$F(1,24) = 11.003$, $p = .003$, $\eta_p^2 = .314$] was found for the visual n-back task. The interaction was not statistically significant. Post-hoc comparisons showed that the mean number of errors was lower for the top than for the standard group ($p = .003$). Also, the mean number of errors was lower in the first session.

For the auditory n-back task, results were the same (main effects of Sessions [$F(3,72) = 58.584$, $p < .001$, $\eta_p^2 = .599$], and main effect of Group, [$F(1,24) = 4.689$, $p = .041$, η_p^2

= .163]). The top group showed fewer errors than the standard group and the mean number of errors in the first session was lower.

Finally, a 2x2x16 ANOVA (Group x Modality x Session) was computed with the mean number of errors in the dual n-back task for each modality. Group (top and standard) and Modality (visual and auditory) were between factors and Session was a within factor (repeated measured factor). Figure 6.11 shows the results.

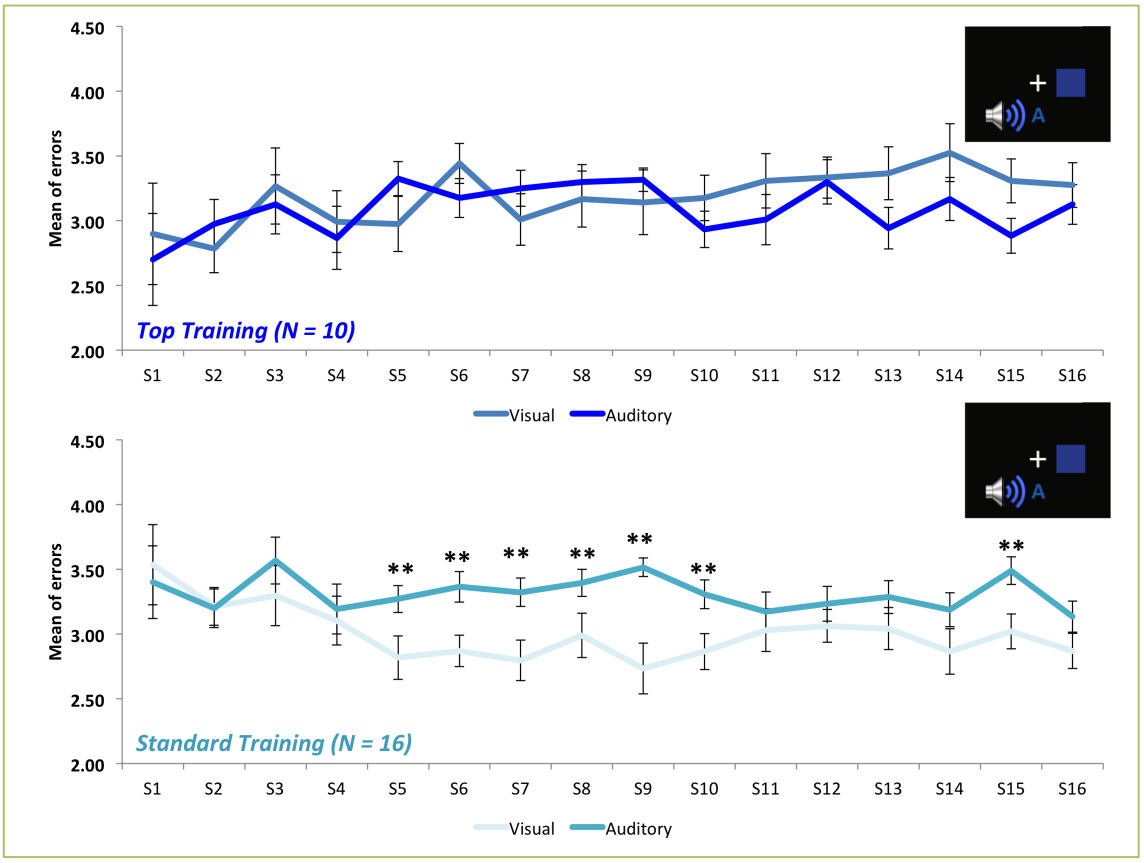


Figure 6.11. Dark blue lines depicts mean of errors for top training participants (N = 10) in dual task. Light blue lines depicts mean of errors for standard training participants (N = 16) in dual task. Auditory modality was represented with a darker colour than visual modality. ** = Significant difference between modalities.

The mean number of errors per block in the top group was between 2.5 and 3.5 across the 16 sessions, while for the standard group these values were between 3.0 and 4.0. No Session effect was found in any group. For the top group, differences between the visual and auditory modalities were not statistically significant, but some statistically significant differences were found in the standard group (their mean number of errors was higher for the auditory modality).

Finally, motivational differences were tested. No statistically significant group differences were found (see results in Supplementary Table 3). Nevertheless, effect sizes revealed moderate group differences in the perceived difficulty level (d = -.69) and perceived challenge of the training program (d = -.80); the standard group found the task was more difficult and challenging than the top group.

6.4.3. *Transfer effects.*

Scores at the construct level for the training (top and standard) and control groups are shown in Figure 6.12.

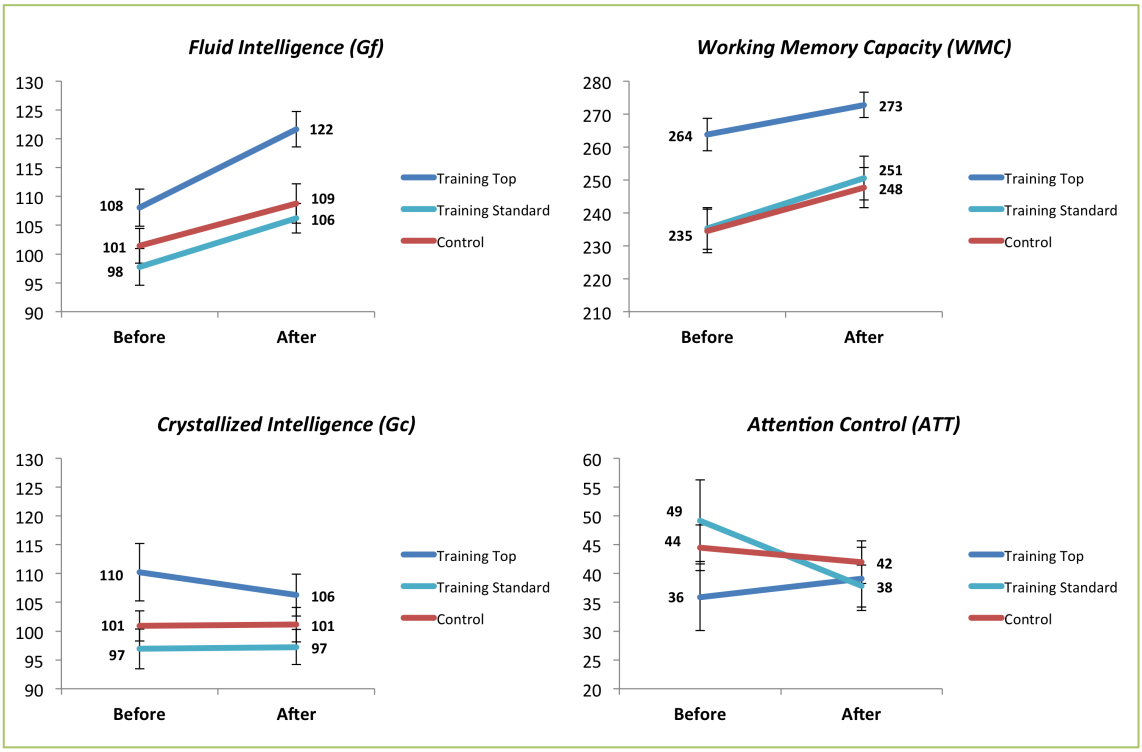


Figure 6.12. Scores before and after n-back training program for the top performance (N = 10), standard performance (N = 16) and control (N = 26) groups at the construct level.

Before training, the standard and control groups showed similar scores in all psychological factors. The top group showed statistically significant greater Gc and WMC scores than the standard and control groups. For Gf, there was a statistically significant difference between the top and standard groups. After training, the group differences remain for WMC and Gf.

Standardized changes [(after − before)/ SD before] (Jaeggi et al., 2011) were computed at the construct and measure levels. Scores were submitted to ANCOVA analyses, where the dependent variable was the standardized score, the independent variable was group and the covariate was the score at the pretest (before training). A *p* level of .05 (one-tailed) was considered (Jaeggi et al., 2008). Results for each cognitive construct are shown in Figure 6.13. ATT scores were reversed for clarity purposes.

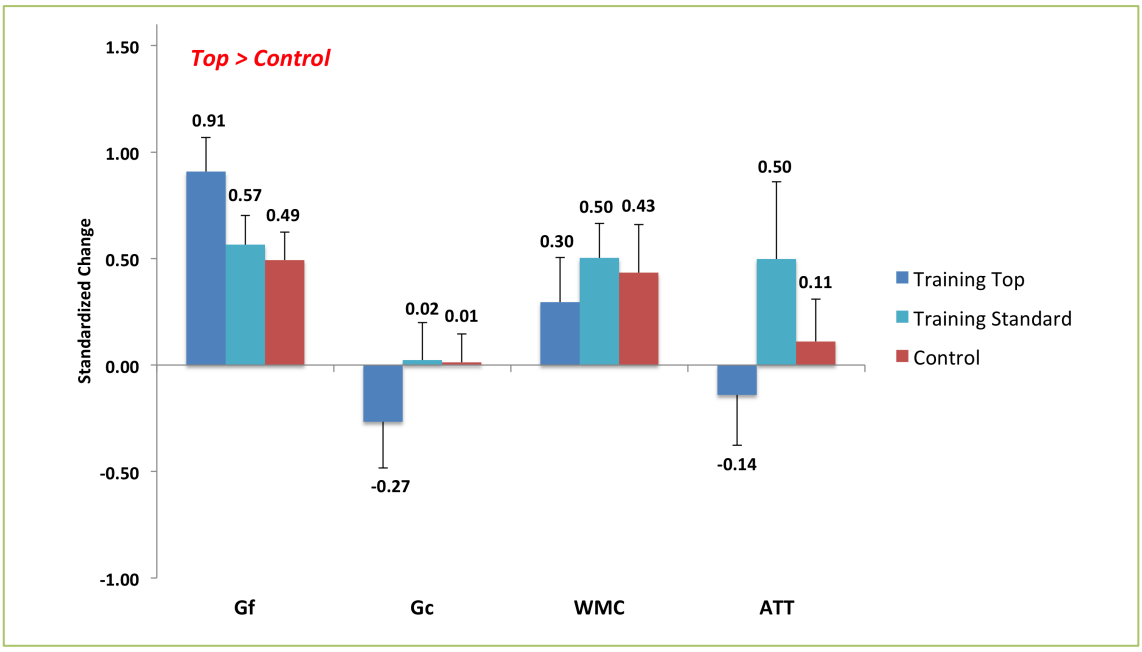


Figure 6.13. Standardized change [(after − before)/ SD before] of the top, standard, and control groups in the assessed psychological constructs. Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity, ATT = attention control.

The standardized Gf change in the top group was higher than in the control group, $F(2,48) = 2.541$, $p = .044$, $\eta_p^2 = .096$ (.91 vs .49). This result is consistent with the findings reported by Jaeggi et al. (2011) with a group of children.

Group differences for the change in Gc, WMC, and ATT were not statistically significant, Gc, $F(2,48) = .109$, $p = .448$, $\eta_p^2 = .005$, WMC, $F(2,48) = .774$, $p = .233$, $\eta_p^2 = .031$, ATT, $F(2,48) = .395$, $p = .338$, $\eta_p^2 = .016$.

Regarding the measure level, descriptive statistics are shown in Supplementary Table 4a (top) and 4b (standard). Figure 6.14 shows these results.

Interestingly, changes in Gf measures were always higher in the top group. However, the difference was only significant for DAT-AR, $F(2,48) = 2.551$, $p = .044$, $\eta_p^2 = .096$,

top > standard. Note that this test showed the highest factor loading (.74) on Gf (see Figure 5.6). Group differences in the RAPM and PMA-R were not statistically significant, RAPM, $F(2,48) = 1.727$, $p = .094$, $\eta_p^2 = .067$, PMA-R, $F(2,48) = 1.525$, $p = .114$, $\eta_p^2 = .060$.

Null group differences were found for the Gc measures: $F(2,48) = .673$, $p = .257$, $\eta_p^2 = .027$ (DAT-VR), $F(2,48) = .841$, $p = .218$, $\eta_p^2 = .034$ (DAT-NR) and $F(2,48) = .410$, $p = .333$, $\eta_p^2 = .017$ (PMA-V).

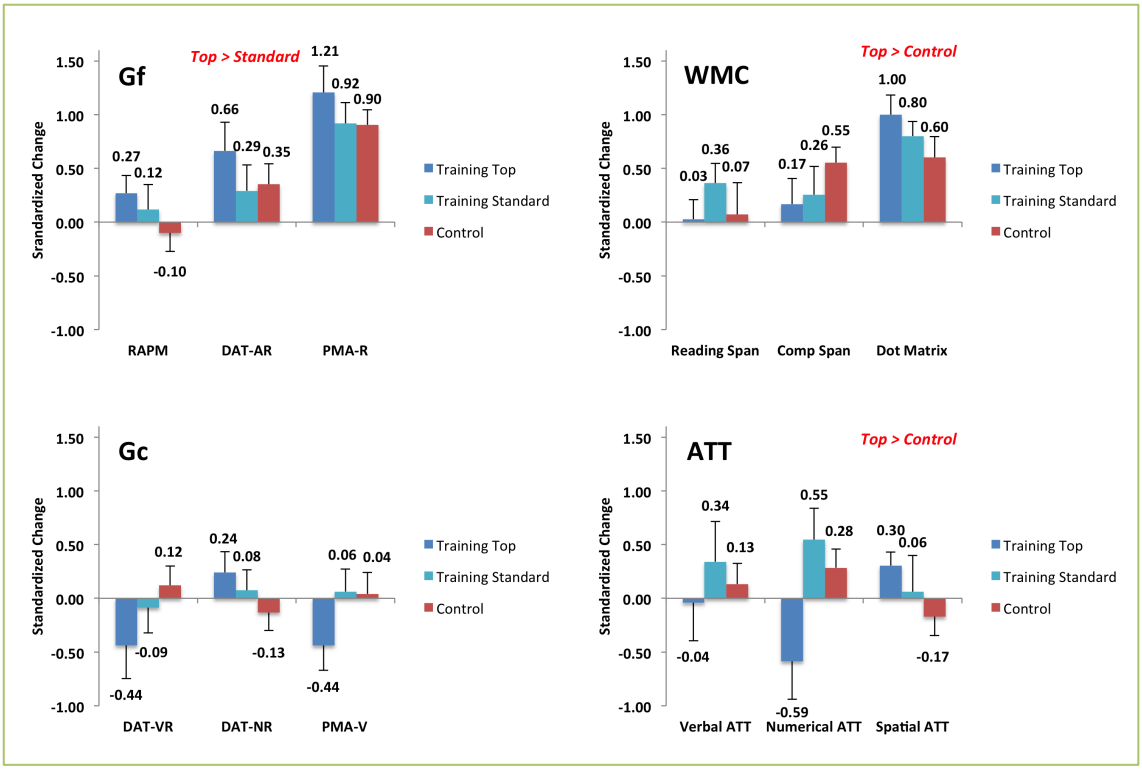


Figure 6.14. Standardized change [(after − before)/ SD before] at the measures level in the top, standard and control groups: Gf = fluid intelligence (RAPM = Raven Advanced Progressive Matrices Test, DAT-AR = abstract reasoning, PMA-R = inductive reasoning); Gc = crystallized intelligence (DAT-VR = verbal reasoning, DAT-NR = numerical reasoning, PMA-V = vocabulary); WMC = working memory capacity; ATT = attention control (verbal ATT = vowel–consonant, numerical ATT = odd–even, spatial ATT = right–left).

Regarding the WMC measures, the difference between top and control groups was significant in dot-matrix task, $F(2,48) = 2.208$, $p = .049$, $\eta_p^2 = .91$. There were no statistically significant group differences for reading span and computation span: $F(2,48) = .715$, $p = .237$, $\eta_p^2 = .032$ (reading span) and $F(2,48) = 1.184$, $p = .157$, $\eta_p^2 = .047$ (computation span).

Finally, statistically significant differences were found in the spatial ATT task, $F(2,48) = 2.827$, $p = .034$, $\eta_p^2 = .105$. The top group outperformed the control group. There were no statistically significant group differences in the verbal and numerical ATT tasks: Verbal, $F(2,48) = .849$, $p = .217$, $\eta_p^2 = .035$, Numerical, $F(2,48) = 1.263$, $p = .292$, $\eta_p^2 = .050$.

6.5. Discussion.

The key finding was that the large improvements observed across the challenging adaptive cognitive training program based on the n-back task (Figure 6.1) did not evoke statistically significant greater changes than those observed for a passive control group on a set of cognitive factors (fluid-abstract intelligence, crystallized intelligence, working memory capacity, and attention). Therefore, the main hypothesis was rejected.

Nevertheless, individual differences in training performance revealed that participants with the greatest improvements (top performers) also showed statistically significant ($p = .044$) improvements in fluid intelligence at the construct level (Figure 6.9). It is noteworthy that top performers failed to show larger improvements in working memory capacity. Therefore, this group had statistically significant changes in a far-transfer construct (Gf) but not in the near-transfer construct (WMC) which suggests that the cognitive processes involved in the training may not be closely associated with the working memory system (Jaeggi et al., 2010a,b; Kane et al., 2007; Schmiedek et al., 2009).

6.5.1. *N-back training.*

Firstly, participants showed increments on the training program equivalent to those observed in the study by Jaeggi et al. (2008) and greater than those reported by Chooi & Thompson (2012) and Redick et al. (2013). Thus, the training program was successful in achieving the required results.

Secondly, performance across the training sessions correlated with fluid intelligence (Gf), crystallized intelligence (Gc) and working memory capacity (WMC), but not with attention control (ATT) and personality traits and. Note that:

(a) The n-back task was designed as a measure of WMC (Conway et al., 2005), and previous research has shown that this task can be considered a proper measure of individual differences in Gf (Jaeggi et al., 2010b). Therefore, the observed correlations are consistent with the available evidence.

(b) The stability of the correlation values across sessions supports the adaptive nature of the training program, meaning that participants remained engaged in fulfilling the requirement of controlled processing. The greater transfer effects observed for the top performers are coherent with this idea.

Thirdly, top and standard performers were identified according to their n-back level achieved at the end of the training program. Nevertheless, their performance levels were sharply different from the outset. Top performers showed less numbers of errors and their errors on the dual n-back task were evenly distributed. In contrast, standard performers showed more errors in the auditory condition, suggesting their substantial difficulties in copying with the simultaneous requirements. Jaeggi et al. (2010a) showed that auditory conditions are more sensitive to increased difficulty levels (longer response latencies and more errors as the levels of "n" increase) (see also Jaeggi et al., 2010b, 2014).

Finally, with respect to motivation (Tapia & de la Red Fadrique, 2007), participants showed proper levels across the entire program (scores close to 7 or higher in a rating scale from very low -0- to very high -10-). No differences between top and standard performers were found regarding involvement with the task, expectations for future achievement or in the general evaluation of the program. However, top performers found the task less difficult than the standard performers, which reinforces the view that the task should be challenging at an individual level (Jaeggi et al., 2011; Ryan, & Deci, 2000).

6.5.2. *Transfer effects.*

The analyzed transfer effects were consistent with previous reports showing a lack of changes at the construct level (Chooi & Thompson, 2012; Redick et al., 2012). Chooi & Thompson (2012) assessed intelligence changes after completing the adaptive n-back dual task. Verbal intelligence, perceptual intelligence, and mental rotation were the

measured constructs. The training, active control, and passive control groups showed decrements at the post-test level in verbal intelligence and perceptual intelligence. No changes were found for the RAPM test (Gf), the Mill Hill test (Gc), and the operation span task (working memory) for the three groups. Redick et al. (2012) achieved similar conclusions after the assessment of fluid intelligence, crystallized intelligence, working memory, multitasking, and processing speed. They failed to find any difference at the construct level and at the measures levels in their training, active control, and passive control groups. Furthermore, Redick et al. (2012) reported a lack of significant correlations between average n-back performance and the measured psychological constructs. Together with the low average n-back performance achieved by their trained participants, reservations can be raised regarding the straightforward comparison of their findings and those observed in the present study.

Nevertheless, a statistically significant difference in Gf was found here for individuals showing the greatest improvements in the training program (top performers). Jaeggi et al. (2011) found similar results studying children: only participants showing large improvements on the training task obtained remarkable improvements in fluid intelligence. Contrary to our results, Gf differences at baseline between children with large and small Gf improvements were not statistically significant. Consistently with our results, they suggested that the critical point to a satisfactory Gf improvement after training is the degree of improvement across sessions.

A close inspection of the findings reported in the present study revealed several noteworthy issues.

First, even when the difference between the training and control group was not statistically significant, the training group showed an increment in Gf twice as big as that observed for the control group. This gap increases when top performers are compared with the control group. At the measure level, the training group showed greater changes in the RAPM and DAT-AR, and, again this was especially true for top performers. A post-hoc power analysis (G*Power software; Faul, Erdfelder, Lang, & Buchner, 2007) indicated that the power was .84 for detecting a large effect size (f = .40) and .45 for a medium (f = .25) effect size. Therefore, only big effects are detected with an appropriate power value.

Second, Shipstead et al., (2012) raised reasonable doubts regarding the use of passive control groups in these cognitive training studies. However, the results reported here with respect to crystallized intelligence suggest that the type of factors enumerated by these researchers (Hawthorne effect, etc.) did not operate in the present study. The standardized changes for the control group in the Gc construct and in their specific measures were parallel to those that were observed for the training group. This is also consistent with the results reported by Chooi & Thompson (2012) and Redick et al. (2012). Finally, note that these types of factors are much less relevant than generally assumed (Kompier, 2006). In fact, recent historical analyses have questioned the data on which the conclusion of the Hawthorne effect was based (Izawa, French, & Hedge, 2011). The division between top and standard performance fail to add anything new in this regard.

Third, Jaeggi et al. (2008) failed to find changes in a working memory measure (reading span) after the application of the cognitive training on the adaptive dual n-back (although, interestingly, they found significant changes for digit span, a pure short-term memory measure). The same lack of change was observed by Chooi & Thompson (2012) and Jaeggi et al. (2010) for the operation span task, as well as by Redick et al. (2012) for the symmetry and running span tasks. Jaeggi et al., (2010a,b) argued that the correlation between the n-back task and WMC complex tasks was low, indicating that they tap related but separable aspects of WM capacity (Jaeggi et al., 2010a,b; Kane et al., 2007; Schmiedek et al., 2009). Here, we have shown that, at the construct level, the standardized improvements were almost the same for the training and control groups. However, two out of three working memory measures showed statistically significant differences between groups. Dot matrix and reading span improvements were substantially higher for the training than for the control group. This result is reversed for the computation span task. Averaging the two types of results produced a null difference at the construct level for working memory capacity, which reinforces the caution note regarding construct heterogeneity (Jaeggi et al., 2010a,b; Schmiedek et al., 2009).

Finally, the general findings for attention were similar to those found for crystallized intelligence: there was a very small standardized change for both groups at the construct level and this was also observed for the specific attention measures. The exception was

for spatial attention, for which the training group showed a standardized change significantly greater than that observed for the control group. This improvement in spatial attention was found for top performers only.

Taken together, the results for the twelve measures administered before and after training suggest that the cognitive intervention used here may enhance visuo-spatial processing (also consistent with Jaeggi et al., 2014). The visuo-spatial fluid measures (RAPM and abstract reasoning - DAT-AR), along with spatial working memory (dot matrix), and spatial attention (Simon task) showed the greatest difference between the training and control groups, favouring the former. This observation is reinforced by the negative results found for the crystallized-verbal measures, computation span (working memory), and the verbal and numerical attention control tasks. The reading span task seemed like an exception to this general pattern. However, it should be noted that there is a clear spatial requirement for this working memory task, since participants must recall the displayed letters (secondary task) according to their 'position' in the alphabet ignoring their serial order in the sequence. Furthermore, the auditory n-back condition was based on the updating of the set of letters and this might have had some positive specific impact here. In fact, the auditory modality was more difficult (higher errors) for the standard group at increased difficulty levels.

The meta-analysis reported by Melby-Lervåg and Hulme (2012) supports the positive result for these visuospatial processing skills. These researchers analyzed twenty-three studies finding reliable short-term and specific increments in working memory skills after cognitive training. Note also that the meta-analyses published by Hindin and Zelinski (2012) and Uttal et al. (2013) found small/medium positive effect sizes for cognitive training in terms of improvement in non-trained domains. One possible explanation for this phenomenon is that people generally have less practice with spatial than with verbal tasks, and thus may have more room for improvement in spatial tasks (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001).

If adaptive working memory/short-term memory training promotes skills relevant for the reliable temporary storage of relevant information, then fluid intelligence (far-transfer) and working memory capacity scores (near-transfer) will be higher for a trained group rather than for a control group. Jaeggi et al. (2014) suggested that executive updating processes, inhibition and interference resolution may support the

relationships among these constructs. In fact, Martínez et al., (2011) demonstrated a near-perfect correlation, at the latent variable level, among short-term memory, executive updating, working memory, and fluid intelligence.

The main conclusion is that the adaptive cognitive training based on the n-back task might increase performance in fluid intelligence at the construct level when high training performance levels are achieved (top performers). Nevertheless, post-hoc analyses performed at the measure level suggest that further research is required to determine whether the administered cognitive training may enhance visuospatial processing skills, as well as the resolution of interference in spatial tasks (Jaeggi et al., 2014).

CHAPTER 7: Gray matter responsiveness and n-back training (Q2.1).

7. 1. Introduction.

This chapter is devoted to the analysis of two morphological indices obtained through a Surface-Based Morphometry (SBM) processing pipeline: cortical thickness (CT) and cortical surface area (CSA) (Burgaleta et al., 2014; Colom et al., 2013b; Panizzon et al., 2009; Román et al., 2014; Vuoksimaa et al., 2014). CSA is thought to be driven by the number and spacing of mini-columnar units of cells, whereas CT is expected to reflect the number of neurons per column (neuron density) as well as glial support and dendritic connections (Chance et al., 2008; Lyttelton et al., 2009). T1-weighted images were employed as input data for SBM analyses. As noted above, four subjects (two per group) failed to pass QC (quality control). Thus, the analyzed sample comprised 26 participants per group.

This chapter addresses the second question (Q2), with CT and CSA as structural indices. Therefore, we focused in the study of changes in these indices as a consequence of the n-back training program. Thus, the interaction (group x time) was studied in order to identify regions where changes were different in both groups (Thomas & Baker, 2013a,b). Frontal, parietal, and middle temporal regions are the regions expected to change after the n-back training program (Takeuchi et al., 2011), as they are crucial in supporting the individual differences in intelligence and working memory (Bassett al., 2015; Burgess et al., 2011; Colom et al., 2013b; Jung & Haier, 2007). These regions are believed to be involved in information integration and evaluation of cognitive processes, which are relevant to working memory (Buschman et al., 2011; Hampson et al., 2006; Rottschy et al., 2012; Zou et al., 2013). Furthermore, a reduction in cortical thickness for the control group is expected, since CT has shown a spontaneous decrease with age (Wierenga et al., 2014; Zhou et al., 2015). For CSA, changes during brain development have been less studied and results are less clear (see Burgaleta et al., 2014; Wierenga et al., 2014; Zhou et al., 2015). Thus, we don't have a specific hypothesis for this index. Nonetheless, the relationship between performance on the n-back task and these structural indices are depicted in this chapter.

The chapter is divided in three sections: (a) analysis protocol, (b) results at the regions of interest (ROI) level[1], and (c) discussion.

7. 2. Surface-Based Morphometry: CIVET 2.0 pipeline.

CIVET (version 2.0) is a structural human MRI cortical surface segmentation tool developed at the Montreal Neurological Institute (MNI). CIVET is a fully automated pipeline to extract and co-register cortical surfaces (Ad-Dab'bagh et al., 2006). The pipeline is represented in Figure 7.1.

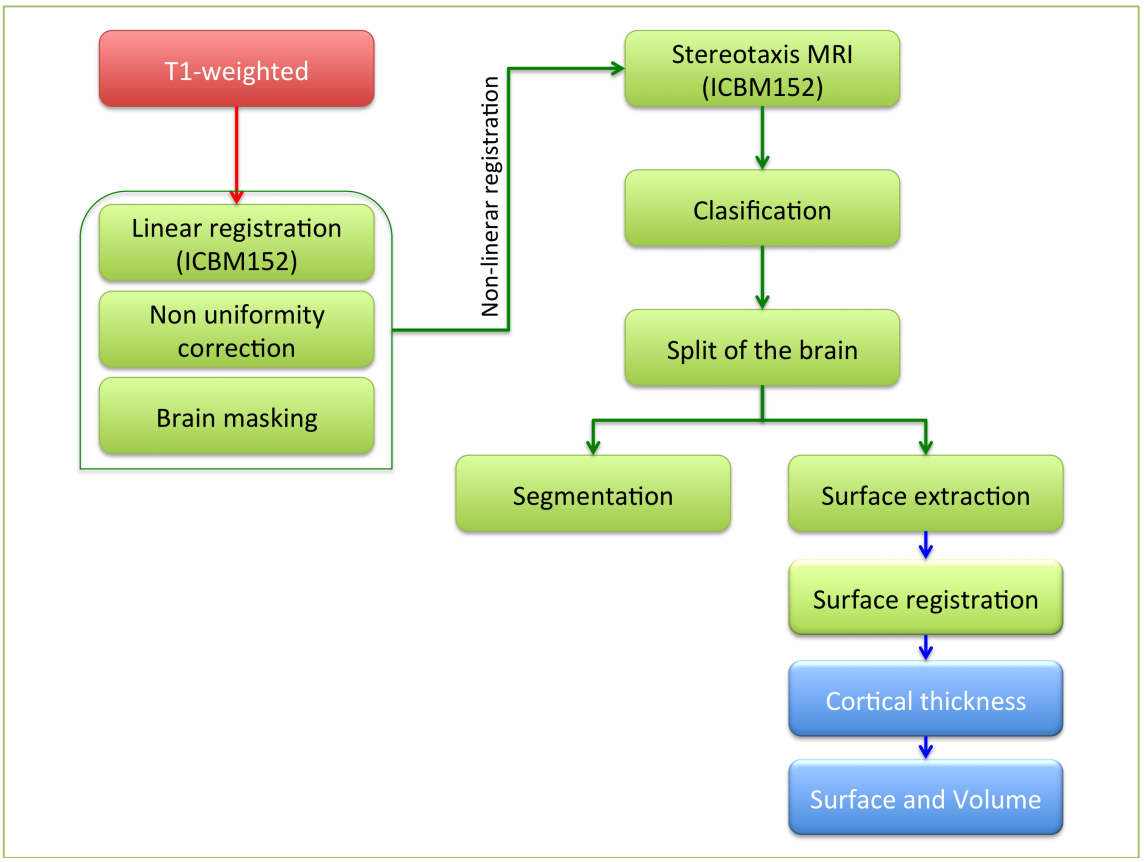


Figure 7.1. CIVET pipeline. Red color represents input data, green color depicts the main steps, and blue symbolizes the outputs.

(1) Linear registration. Involves deformation from native brain scans to a 3D template brain, which defines a common coordinate space for analysis of a given dataset. Each individual MR image is registered to the same template using a linear/affine registration with 9 or 12-parameter (Fonov, Evans, McKinstry, Almli, & Collins, 2009). The template target of registration used here was Average MNI ICBM152 (see Figure 7.2).

[1] Analysis at vertex level can be found in Appendix 6.

A registration of 12 parameters was used for this sample (3 rotations, 3 translations, 3 scaling, and 3 shears from native image to stereotaxic space).

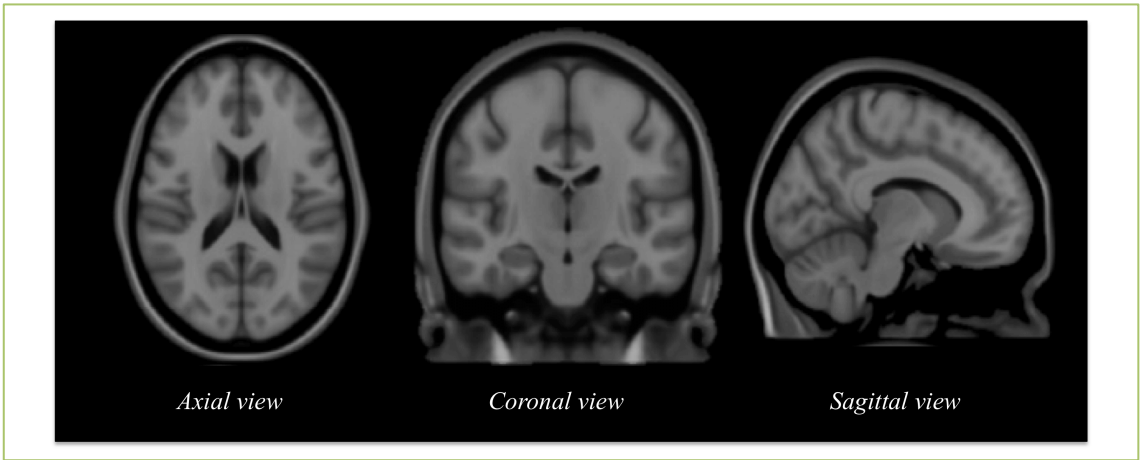


Figure 7.2. The average MNI ICBM 152 represented in different views: axial, coronal and sagittal.

(2) Correction of non-uniformities and brain mask. Images obtained from registration are corrected for radio frequency non-uniformities to delete non-biological artifacts (Sled, Zijdenbos, & Evans, 1998). Afterwards, a brain mask is computed for each individual (Smith, 2002). See Figure 7.3 for an example of both processes. Then, a non-linear transformation (Collins, Neelin, Peters, & Evans, 1994) is computed from each subject in stereotaxic space.

(3) Tissue classification (Tohka, Zijdenbos, & Evans, 2004) into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) (Figure 7.4). The classification comprises two steps: a discrete classification based on tag points, followed by the evaluation of the partial volumes for the tissue classes. The tissue classification can be based on the T1-weighted image only, or on multispectral data (T1-weighted, T2-weighted, Proton Density weighted). Only the T1-weighted was used in our study.

(4) Boundary masks. The brain is divided into left and right hemispheres for surface extraction. The boundary between cortical GM and subcortical WM (referred to as the white matter surface) is firstly identified and a mesh of triangles is built (Kim et al., 2005). Next, the pial surface, or the boundary between the cortical GM and the extra-cortical CSF (referred to as the GM surface) is constructed by expanding outward the WM surface to the CSF. Each default surface (per hemisphere) is composed of 81,920

triangles (polygons) and 40,962 vertices (MacDonald, Kabani, Avis, & Evans, 2000). However, we used high-resolution surface available in CIVET 2.0 with 327,680 polygons and 163,842 vertices per hemisphere. See Figure 7.5 with examples of these steps.

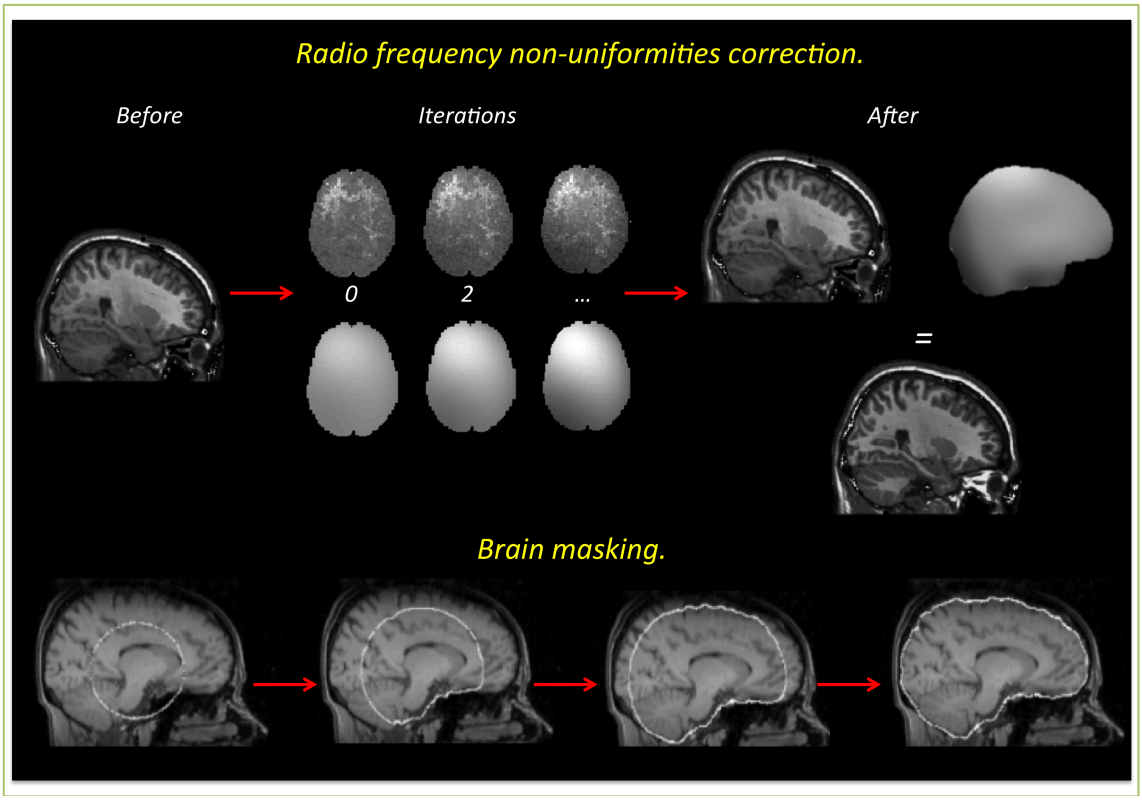


Figure 7.3. Top panel shows the correction of non-uniformity. From left to the right: (a) A slice form MRI image, (b) example of how local and smooth bias field estimates evolve with iterations during correction of the MR volume, (c) The MR volume corrected is obtained by removing the last estimated smooth field. Bottom panel shows example of outputs from the iterative algorithm involved in brain masking process. Figures were adapted from Sled et al. (1998) and Smith (2002).

(5) Computation of cortical thickness and cortical surface area. The surfaces are registered to the MNI ICBM152 surface template (Lyttelton et al., 2007). Cortical thickness is computed by evaluating the distance (in mm; *tlink*) between the original WM and GM surfaces transformed back to the native space of the original MR images, then interpolated onto the surface template (Lerch & Evans, 2005). Cortical surface area was calculated as the sum of the surface area of the 6 triangles linked to each vertex divided by 6. Volume combines cortical thickness (CT) and cortical surface area (CSA). See Figure 7.6 for an example of estimation of CT and CSA.

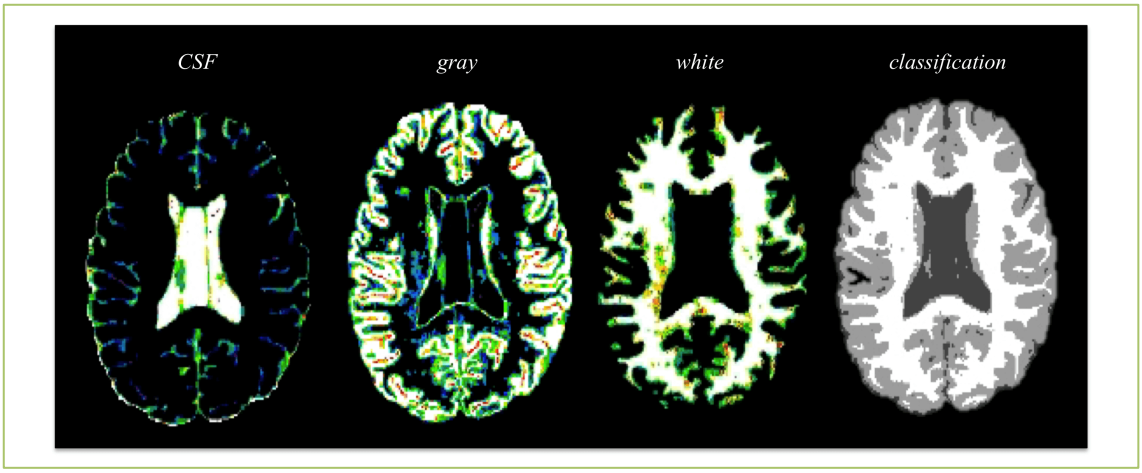


Figure 7.4. Examples of tissue segmentations obtained from CIVET.

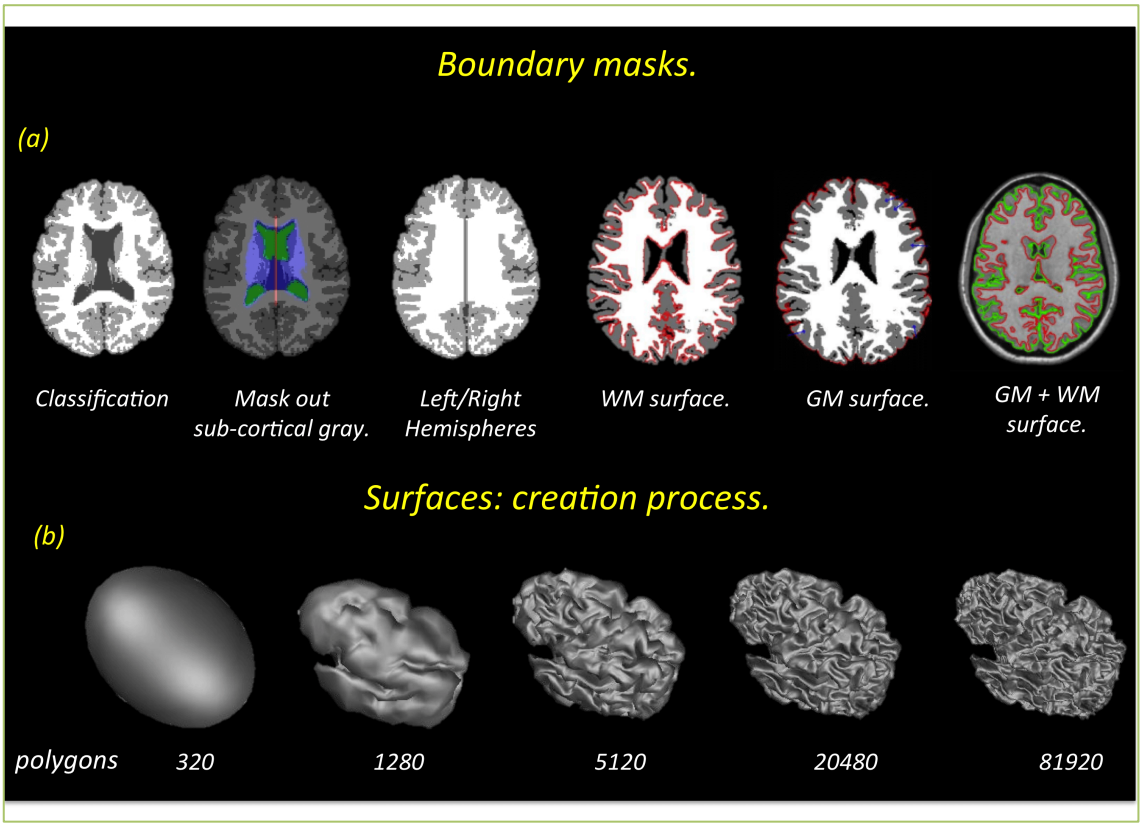


Figure 7.5. In top panel are examples of different boundary masks created by CIVET. Bottom panel shows the creation process of surfaces. WM = white matter; GM = gray matter. Some images were adapted from Kim et al., (2005) and MacDonald et al., (2000).

The quality control (QC) of each participant was performed after processing the 3D-T1 weighted images with the CIVET pipeline. Images were revised for (a) registration, (b) surface extraction, (c) gray-white surface-surface intersections and (d) expansion of the gray surface. One example of the QC for one subject is presented in Appendix 5.

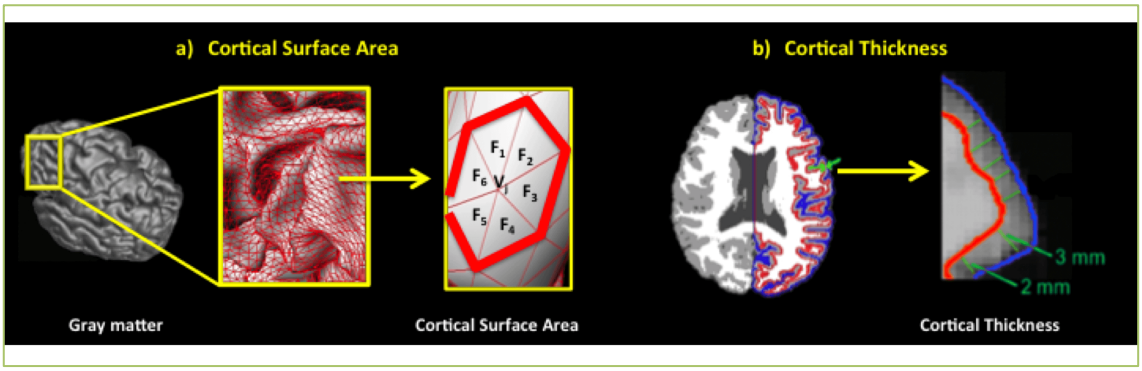


Figure 7.6. Left panel shows an example of the computation of cortical surface area. Right panel shows an example of the estimation of cortical thickness. Figure was adapted from Román et al., (2014).

7. 3. Changes in cortical thickness and cortical surface area (ROIs level).

SBM statistical analyses were computed with SurfStat: (http://www.math.mcgill.ca/keith/surfstat/) created for MATLAB 7 (The Math- Works, Inc.). Findings at vertex level were meager because the results did not pass post-hoc corrections (see Appendix 6). However, the computed effect sizes (*d*) showed several brain regions with values greater than 0.80 (a large effect). Salgado-Pineda et al. (2003) did argue that corrections over the whole brain applied to structural data are too strict, since multiple comparisons were originally designed to analyze functional imaging data.

The cerebral cortex is a highly organized and complex structure divided into anatomically distinct and functionally specialized regions. However, there is no a definitive division because available neuroimaging software is based on different brain parcellations (e.g., *FreeSurfer, BrainSuite, IBASPM, ANIMAL*, etc.). We chose the parcellation noted above because it is data-driven. Therefore, based on Chen et al. (2012, 2013), a set of ROIs, distinguished by their genetic substrate, were selected for both CT and CSA. The main results derived from the Chen et al.'s reports were: (a) CSA is anterior-posterior organized, while CT is dorsal-ventral organized; (b) The genetic patterning of both gray matter indices corresponds to functional specializations and their genetic contributions show very small correlations; (c) the regions/clusters for CSA show genetic proximity within the same brain lobe, whereas regions for CT show genetic relatedness regarding maturational timing (primary vs. association cortex. Therefore, the conclusion obtained by the authors is that neurodevelopmental

mechanisms are different for CT and for CSA. Our study focuses on the analysis of which brain regions, under remarkable genetic control, are sensitive to the completed challenging cognitive training. Also, published reports have massively analyzed volumetric changes (as noted above) and it is known that volume combines CSA and CT, which, as described above, are genetically unrelated. Therefore, the separate analysis of these two gray matter indices would be highly informative

Note that ROIs chosen here are selected from two independent studies to avoid the circularity problem (Thomas & Baker, 2013a). A problem of circularity is found in several studies on this topic where only the interactions in voxel that were significant were analyzed. This approach produces a selection bias that has been demonstrated to produce potentially spurious effects (Baker, Hutchison, & Kanwisher, 2007; Kriegeskorte, Lindquist, Nichols, Poldrack, & Vul, 2010). We decided to analyze the interaction (group x time) in regions genetically relevant and independently of significant results at the vertex level. Figure 7.7 shows the genetic ROIs for CT and Figure 7.8 depicts the ROIs for CSA.

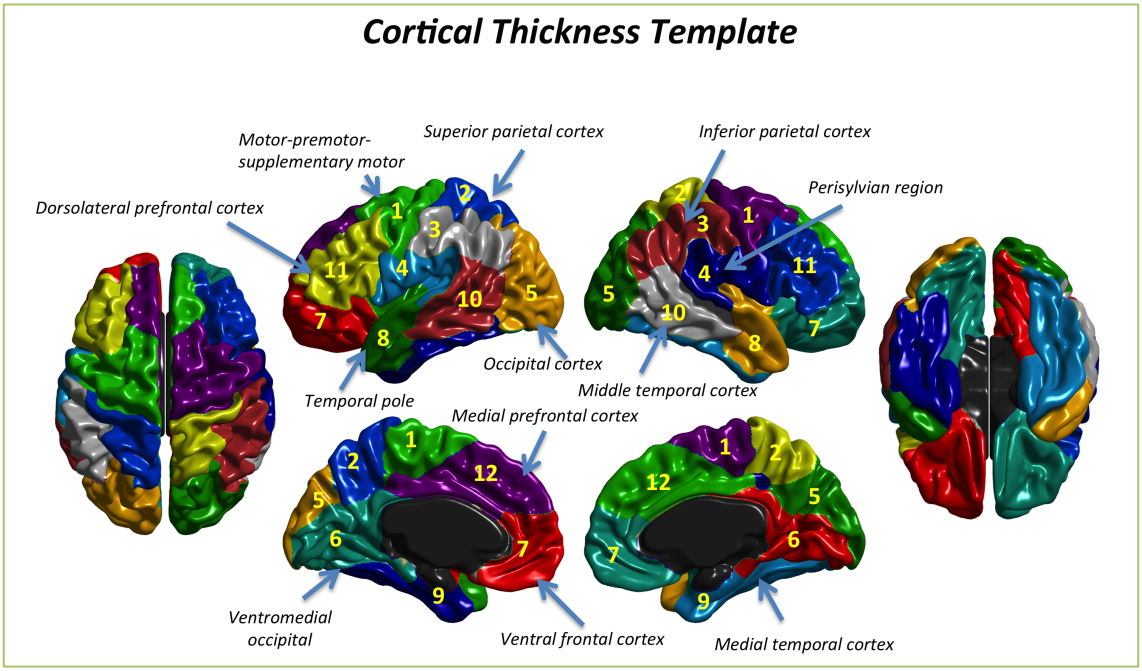


Figure 7.7. Genetic region of interest for cortical thickness.

These ROI analyses were computed using SurfStat. Firstly, potential changes were checked in the 12 brain regions for CSA (Chen et al., 2012) and in the 12 regions for CT (Chen et al., 2013). This required translating the original FreeSurfer coordinates for

each cortical region into MNI space (ICBM152 average surface template). Lindsay B. Lewis at MNI carried out this process. After this conversion, each vertex of MNI space was assigned to a specific ROI using Bayesian theory. This process produced separated regions for each hemisphere, and, therefore, 24 regions were finally retained (see Figure 7.7 for CT and Figure 7.8 for CSA). Finally, mean of CT and total CSA values for each ROI and each participant were computed.

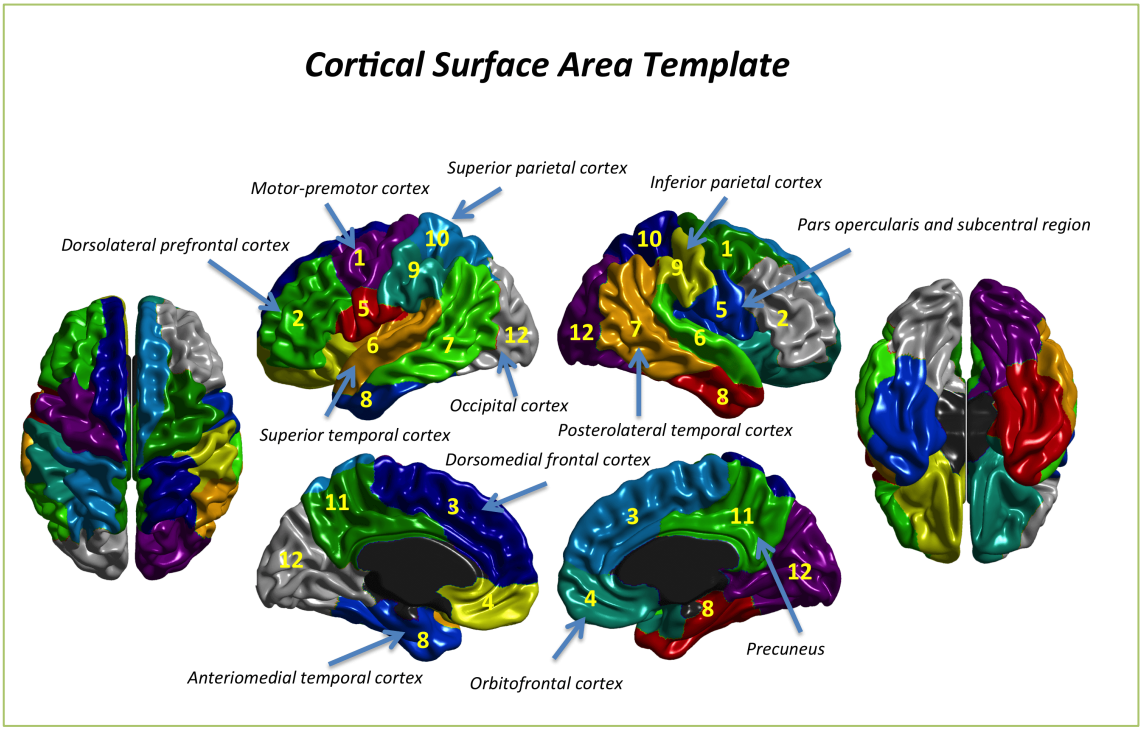


Figure 7.8. Genetic region of interest for cortical surface area.

With the obtained values, we computed standardized changes per subject of the sample using the following formula: (ROI's morphological index after training – ROI's morphological index before training) / SD of ROI's morphological index (before training) (Jaeggi et al., 2011). These standardized changes were submitted to analyses of covariance (ANCOVA) where group was the independent variable, standardized change was the dependent variable for each ROI, and the covariate was the mean for CT or CSA before training for the corresponding ROI. A one-tailed *p* level was considered. Significance level ($\alpha = .05$) was modified according to the number of comparisons and the correlations between ROIs to correct for multiple comparisons using the partial Bonferroni correction, since multiple measurements were computed (24 ROIs per index and per subject. Also, as for vertex level, the following was performed:(a) an analysis of covariations between morphological indices before training and n-back training

performance, (b) an analysis of the group by time interaction, and (c) an analysis of the interactive effects between groups according their n-back level (standard and top) and subjects from control group.

Figure 7.9 shows the mean of correlations for each ROI between (visual and auditory) n-back performance and cortical indices (CT and CSA), since performance in each session was correlated with structural indices and after the mean of all correlations was calculated. Correlations higher than .20[2] were located in bilateral medial temporal cortex, left perysilvian, temporal pole, occipital lobe, right inferior parietal and middle temporal for visual n-back task. Correlations between CT and auditory tasks were found in bilateral medial temporal cortex, occipital and perysilvian, left temporal pole, right inferior parietal, middle temporal, and motor area.

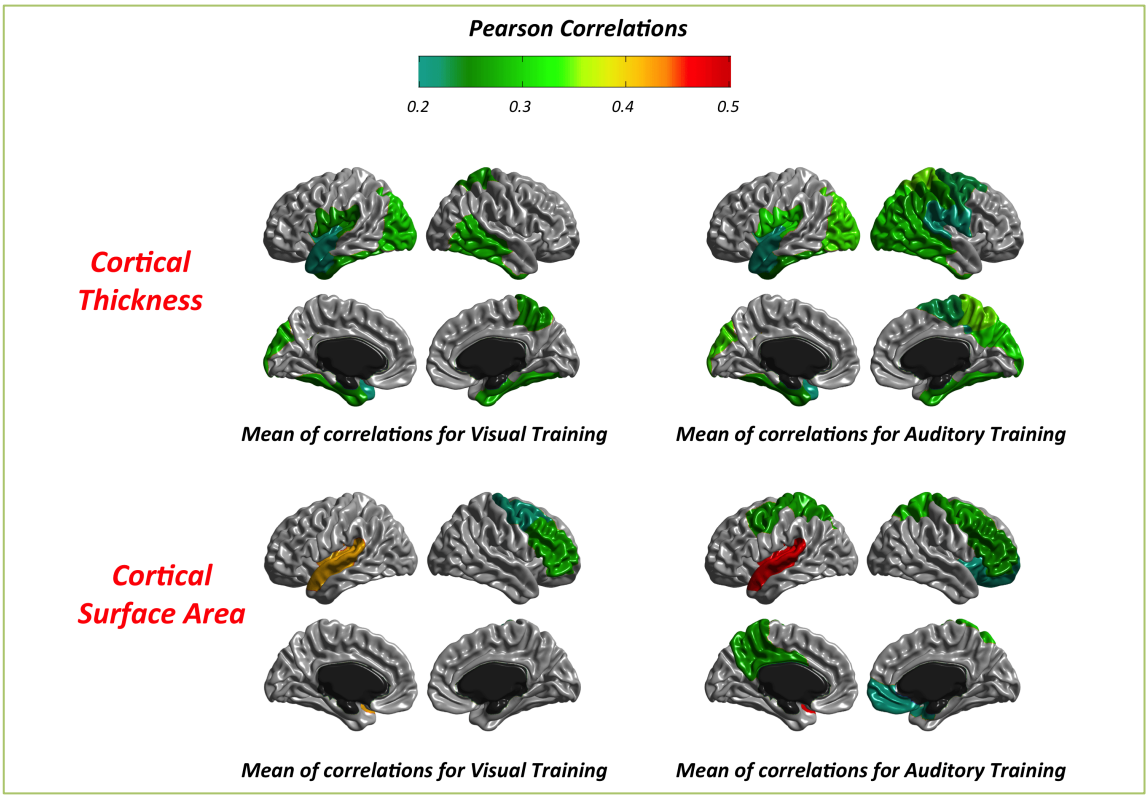


Figure 7.9. Mean of correlations between single n-back training program and genetic-ROIs of cortical thickness and cortical surface area (before training). Only values higher than 0.20 are depicted in the figure.

Results for CSA were located in left superior temporal and right dorsolateral prefrontal and motor area for both tasks. Also, for the auditory task, bilateral inferior parietal, left

[2] Authors such as Shaw et al., (2006), Vuoksimaa et al., (2014) etc, detected that mean of correlations between structural indices and intelligence measures is .20. Thus, we selected this value as cut-off.

precuneus and right orbitofrontal showed values higher than .20. Especially relevant were the correlations between left superior temporal and n-back performance, since the correlation was higher than .40 for the visual task and close to .50 for the auditory task.

Correlations higher than .20 between dual n-back performance and cortical thickness measured before training are depicted in Figure 7.10.

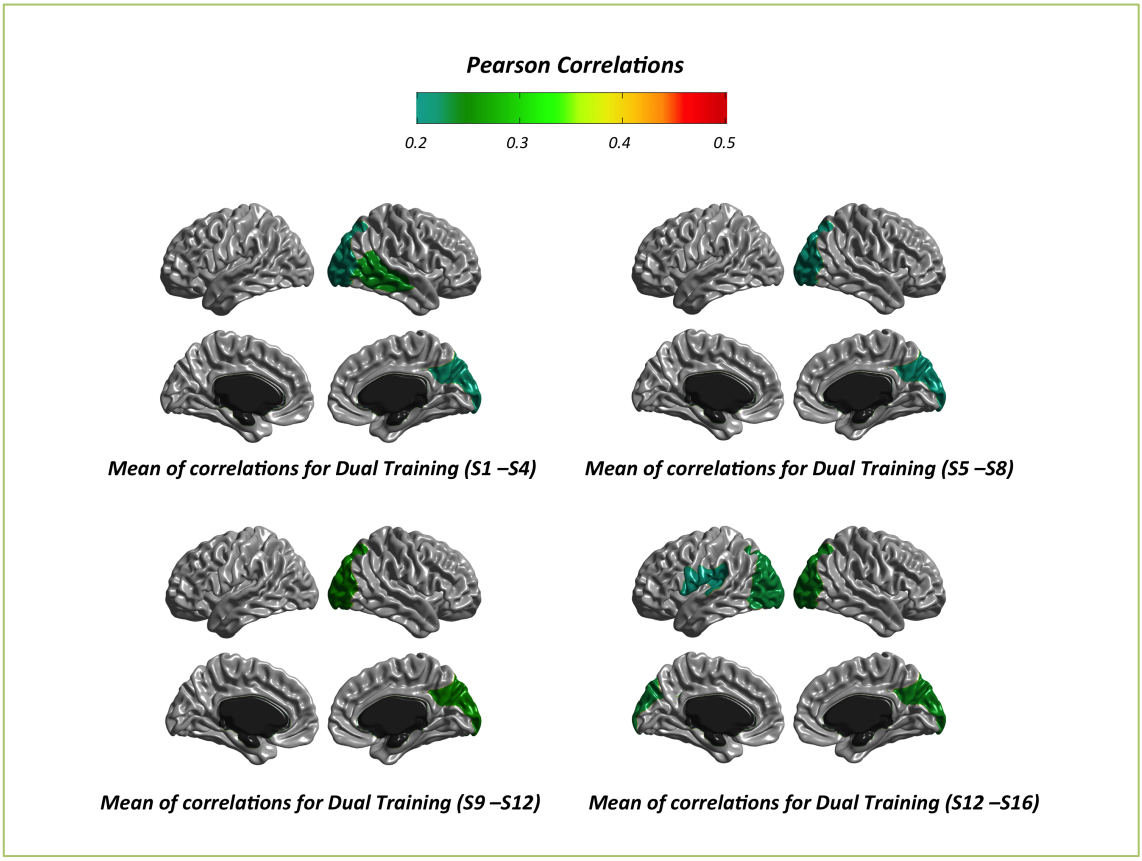


Figure 7.10. Mean of correlation between dual n-back training program and genetic-ROIs of cortical thickness (before training). Only values higher than 0.20 are depicted in the figure.

The right occipital lobe showed *rs* > .20 across all sessions. Also, the right middle temporal and left perysilvian showed *rs* > .20 in the first and last training sessions, respectively.

Figure 7.11 shows the results for CSA. The left superior temporal and right motor area showed correlations higher than 0.20 with performance across all blocks. In the first block, left pars opercularis and precuneus were found as relevant. The precuneus and superior temporal were relevant in the second block. For the third block, bilateral superior parietal and left motor area showed correlations higher than .20. Finally,

correlations between performance in last block and CSA before training were higher than for the remaining blocks. Specifically, bilateral posterolateral temporal, motor, superior parietal and orbitofrontal showed correlations higher than .20. Also, the superior temporal and precuneus were relevant in left hemisphere, along with the dorsomedial frontal and anterior temporal cortex in right hemisphere.

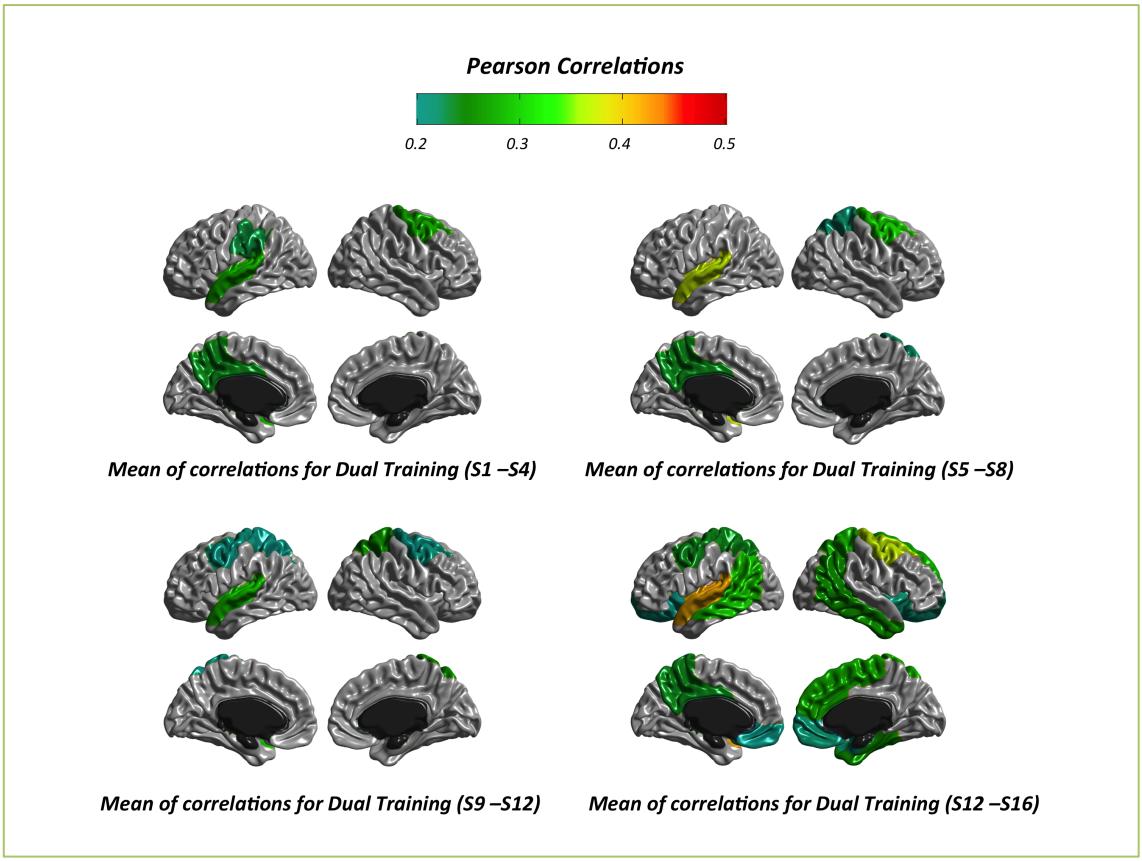


Figure 7.11. Mean of correlation between dual n-back training program and genetic-ROIs of cortical surface area (before training). Only values higher than 0.20 are depicted in the figure.

Afterwards, the interaction time by group was analyzed. Mean CT values for each ROI before and after training are shown in Supplementary Table 5a; results for CSA are shown in Supplementary Table 5b. Standardized changes [(after training – before training) / SD (before training)] were computed for each group and each ROI before and after training.

Figure 7.12 (top panel) shows the standardized change for cortical thickness in the training (left) and control (right) groups. The latter group showed a generalized thinning, whereas the training group hardly showed any change (thinning in the left

middle temporal, frontal and occipital areas, along with thickening in the right temporal region were minor exceptions). The descriptive statistics are shown in Supplementary Table 6a.

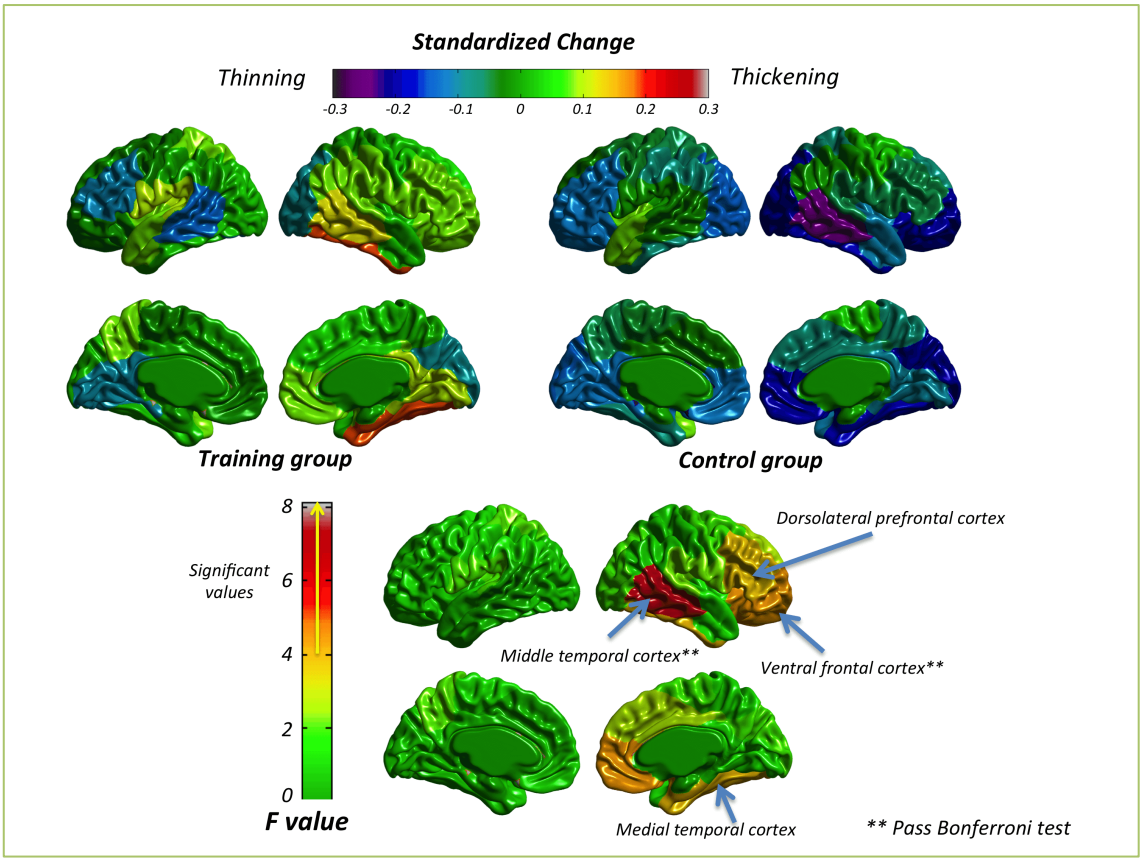


Figure 7.12. Standardized change [(after − before)/ SD at the before] for the training and control groups (top panel) and ANCOVA results (bottom panel) for cortical thickness.

The bottom panel of Figure 7.12 shows the ANCOVA results (*F* value); the specific *F* value and the associated effect (η_p^2) are shown in Supplementary Table 6a. Values greater than 4 were significant. Regions within the right hemisphere showing significant results included the ventral frontal cortex, medial and middle temporal cortex, and dorsolateral prefrontal cortex. However, only the ventral frontal and middle temporal cortex survived the Bonferroni correction ($p < .02$, for $r = 0.75$, number of comparisons = 24). All significant regions had a moderate effect size ($0.06 < \eta_p^2 < .13$; Cohen, 1992). Standardized changes in significant regions showed thinning in the control group (-0.05 to -0.25) and thickening in the training group (0.08 to 0.19).

Figure 7.13 (top panel) shows the standardized change for cortical surface area in the training (left) and control (right) groups (Supplementary Table 6b shows the descriptive results for each group and each ROI). Changes were generally absent for both groups. However, a small contracting effect was found for the control group in the right superior and posterolateral temporal cortex, along with a small expanding effect in the right dorsomedial frontal and anteromedial temporal cortex. For the training group, only a small expanding effect was found in the posterolateral temporal cortex.

ANCOVA results (see bottom panel of Figure 7.13 and Supplementary Table 6b) showed significant values in the right posterolateral temporal cortex and right pars opercularis cortex. Both regions passed the Bonferroni correction ($p < .04$, for $r = 0.98$, number of comparisons = 24) and effect sizes were moderate ($0.06 < \eta_p^2 < .13$; Cohen, 1992). The control group showed a contracting effect of -0.15 (right posterolateral temporal), whereas the training group showed an expanding effect of 0.10 in the same region. For the right pars opercularis, the changes revealed a contracting effect in the control group (-0.09) and a small expanding effect (0.01) in the training group.

Finally, the comparisons between top training performers, standard training performers and controls were computed. Figure 7.14 shows the results for cortical thickness. The top panel shows the standardized changes for three groups, while the bottom panel shows the ANCOVA results. Controls and top performers showed a generalized thinning across the whole brain, although changes in controls were greater. However, standard performers showed the opposite pattern, namely, generalized thickening. Supplementary Table 7a shows the standardized change of cortical thickness and ANCOVA results for each ROI.

ANCOVA analyses showed statistically significant differences in right frontal (dorsolateral prefrontal and ventral frontal) and right temporal (middle temporal and medial temporal) regions. The increase observed for standard performers was higher than for controls, although only results for the right middle temporal passed the Bonferroni correction. No significant differences were found between top performers and the remaining groups.

Figure 7.15 shows the findings for cortical surface area. Standardized changes for controls were located mainly in right temporal lobe (superior and middle) and right pars

opercularis (contracting). Top performers showed mixed results: (a) contracting in bilateral anteromedial temporal, and (b) expanding in left posterolateral temporal. Finally, standard performers also showed mixed results: (a) contracting in left middle temporal (superior and posterolateral) and (b) expanding in right posterolateral temporal, left anteromedial temporal and right orbitofrontal cortex. Supplementary Table 7b shows the standardized change and ANCOVA results for CSA computed on each ROI.

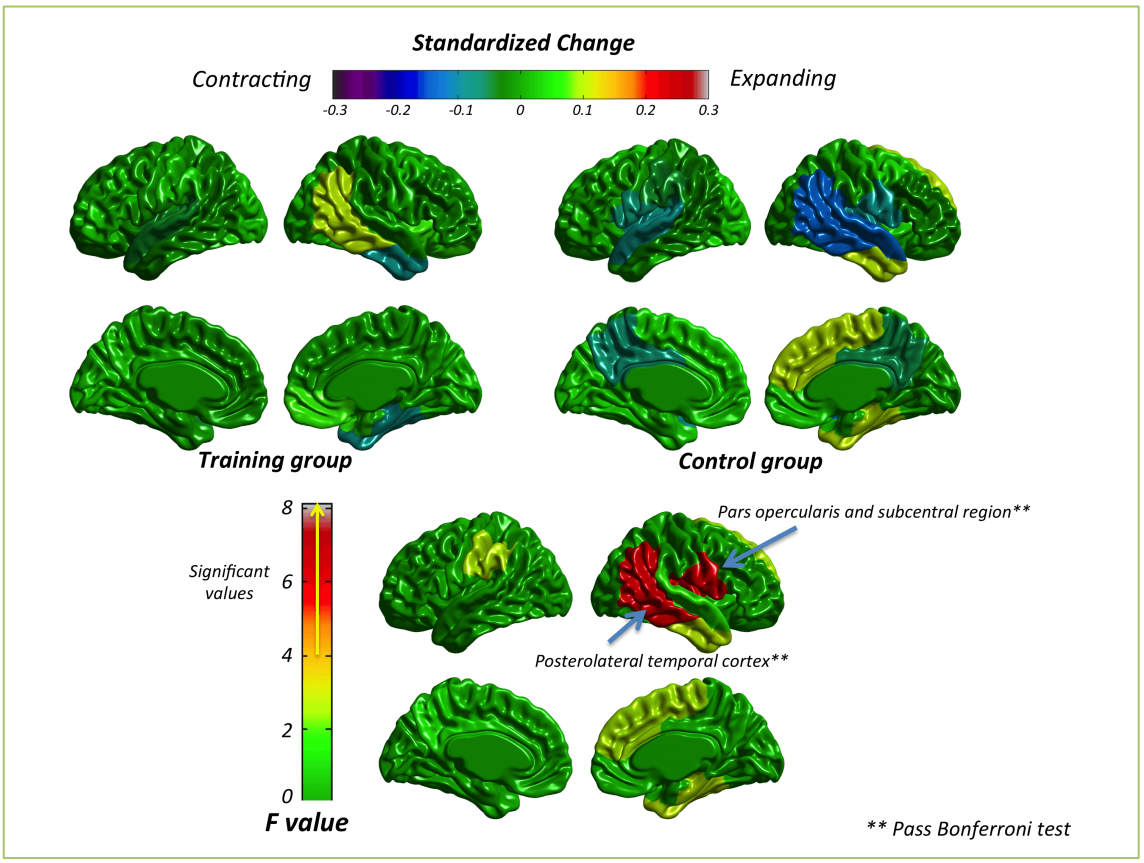


Figure 7.13. Standardized change [(after − before)/ SD at the before] for the training and control groups (top panel) and ANCOVA results (bottom panel) for cortical surface area.

In short, results at ROIs level were: (a) correlations between n-back performance and CT were mainly located in posterior regions, while for CSA the correlations were in anterior regions; (b) the standard training group showed a thickening process compared to standard and control groups in right middle temporal, dorsolateral prefrontal cortex, ventral frontal cortex and medial temporal cortex; (c) for CSA, the outcomes of the comparison were varied: higher increment in the standard training group than in the

control (right posterolateral temporal) and top (right medial temporal) groups, and a difference between top training and control groups located in pars opercularis.

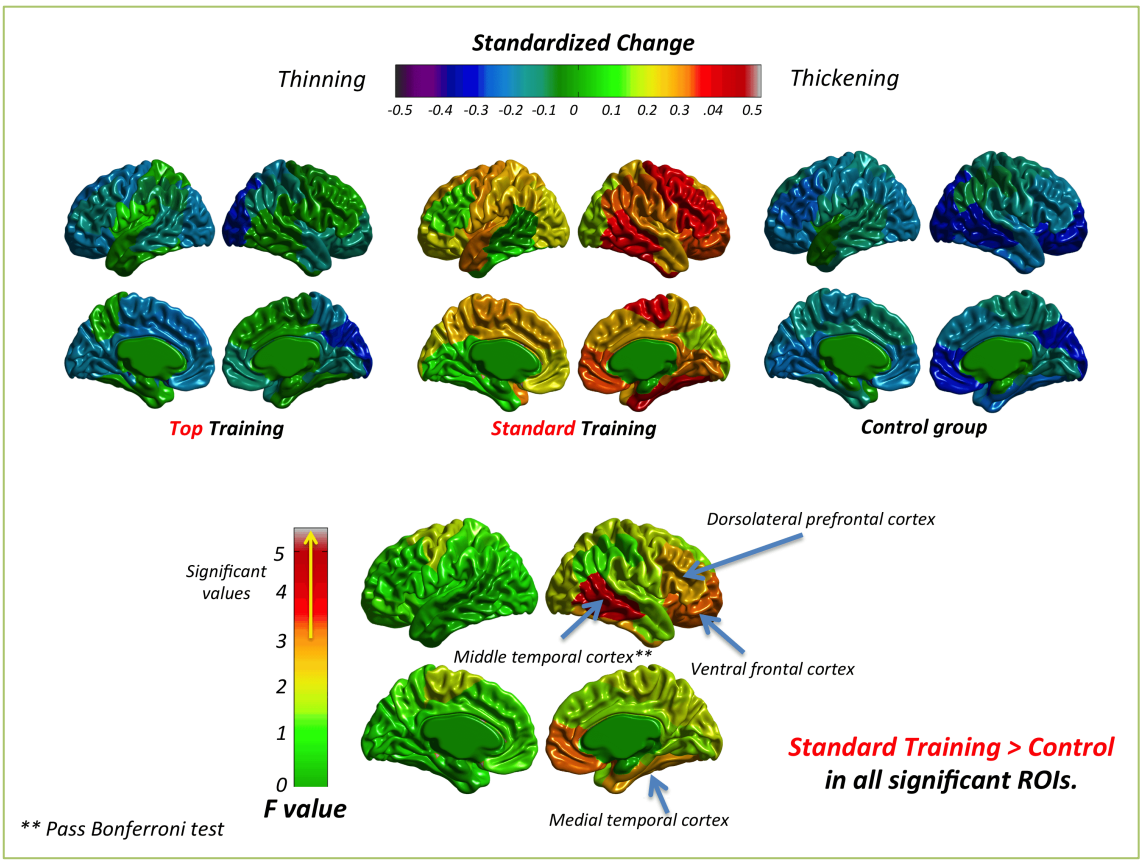


Figure 7.14. Standardized change [(after − before)/ SD at the before] for training (top & standard performance) and control group (top panel) and ANCOVA results (bottom panel) in cortical thickness analysis.

7. 4. Discussion

Here we have analyzed gray matter changes after completing a challenging adaptive cognitive training requiring working memory related skills along with their correlations with cognitive performance. The morphological measures considered were cortical thickness (CT) and cortical surface area (CSA), computed after applying Surface-Based Morphometry (SBM). The analyses were calculated at the vertex and ROIs level. The interaction between group (training vs. control) and time (before vs. after training) was computed, as recommended (Thomas & Baker, 2013a,b). We hypothesized that anterior frontal, parietal and middle temporal regions would show changes after cognitive training (Buschman, et al., 2011; Colom, et al., 2007; Hampson, et al., 2006; Jung & Haier, 2007; Takeuchi et al., 2011).

Note that the time interval considered in our study appears to be enough in order to maximize the probability of evoking detectable structural changes in the brain, since most authors accept that three months is a proper time interval for effectively evoking changes in brain structure (Buschkuehl et al., 2012). There are several examples that support this affirmation. For example, Sagi et al. (2012) found grey matter changes after a short-term period of time. Similar results were found for May et al., (2007), which showed macroscopic gray matter changes after seven days of delivering transcranial magnetic stimulation towards the superior temporal cortex, concluding that structural plasticity in humans can be observed within very short periods of time [see Ishibashi et al., (2006) with changes in white matter structures]. The main findings are discussed next.

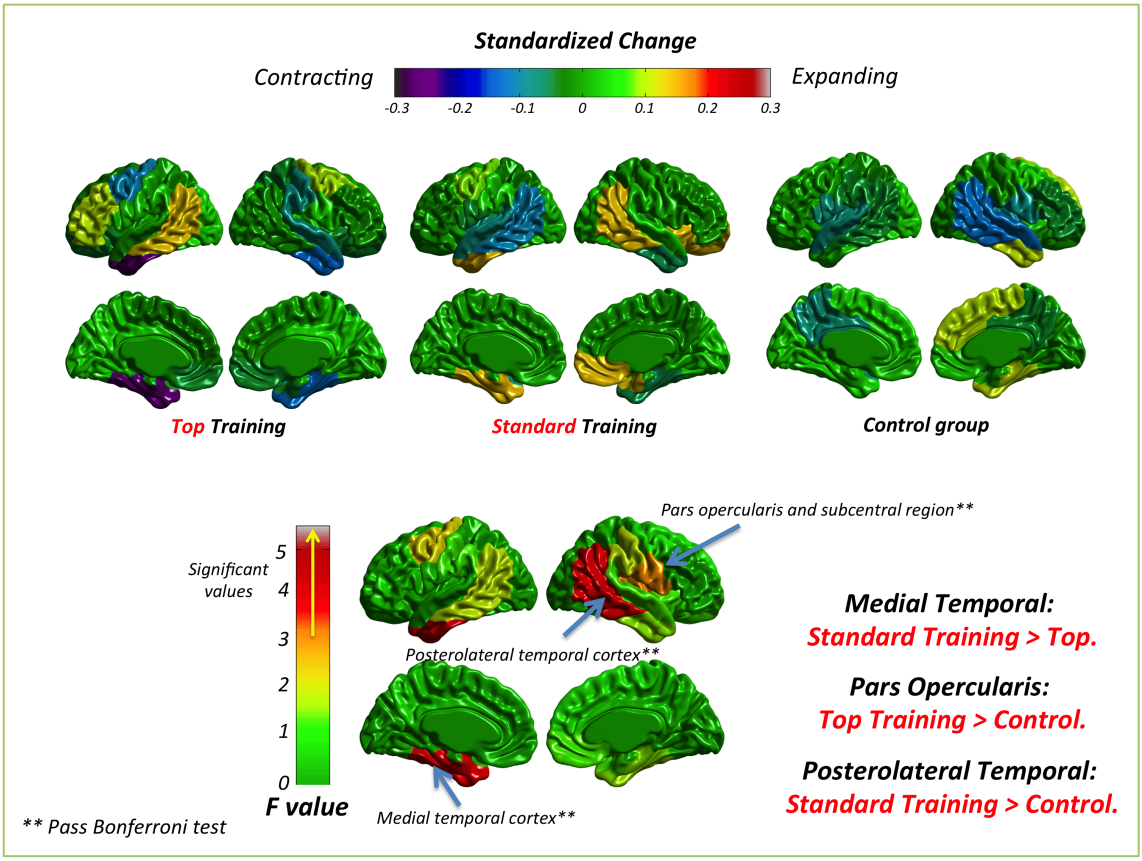


Figure 7.15. Standardized change [(after − before)/ SD at the before] for training (top & standard performance) and control group (top panel) and ANCOVA results (bottom panel) in cortical surface area analysis.

7.4.1. *Correlations with n-back performance.*

The n-back task is a popular measure in functional neuroimaging studies (Conway et al., 2005), but the number of studies reporting structural correlates is very small. Barbey et al., (2014a) showed that low n-back performance is associated with damage to a network comprising regions within right ventrolateral prefrontal cortex, right inferior parietal cortex, and right middle temporal gyrus. Tsuchida and Fellows (2009) found that left lateral and medial prefrontal cortex are critical in letter 2-back performance. Also, damage to dorsal anterior cingulate cortex and adjacent dorsal fronto-medial cortex led to a pattern of impairment marked by high false alarm rates.

The gray matter results reported here were weak at the vertex and ROI levels. Nonetheless, for CSA, anterior regions (dorsal and ventromedial frontal cortex, motor cortex, precuneus, superior-middle temporal and part of inferior parietal cortex) were related with n-back performance, while correlations with CT were located mainly in posterior regions (occipital lobe, middle and inferior temporal cortex and parietal lobe). CSA results were consistent to some degree with previous findings (Barbey et al., 2014a; Tsuchida & Fellows, 2009), perhaps because CSA is strongly related with gray matter volume whereas CT is not (Colom et al., 2013b; Román et al., 2014; Vuoksimaa et al., 2014).

7.4.2. *Changes in Cortical thickness.*

Here, thinning processes were observed in the control group (Figure 7.12), consistent with findings reported in brain developmental studies (Shaw et al., 2006; Zhou et al., 2015). In this regard, Zhou et al., (2015) found, across a ten-year period (from 10 to 20 years), a generalized CT decrease (92%) or no changes (8%). This thinning process is usually attributed to synaptic pruning (Huttenlocher & Dabholkar, 1997), ongoing myelination, and dendritic arborization (Paus, 2010; Sur & Rubenstein, 2005).

The training group did not show CT changes, although small variations were observed in some ROIs: (a) thickening in the right middle and medial temporal and left pars opercularis, and (b) thinning in left middle temporal, dorsolateral prefrontal frontal and ventromedial cortex, and right occipital cortex. This preservation/thickening of cortical thickness for the training group might be a response to training, consistent with previous findings (Engvig et al., 2010; Haier et al., 2009). However, when the training group was

divided according to performance levels, results showed a) similar thinning in top performers and controls, along with b) thickening in standard performers (see Figure 7.14). This can be seen as consistent with the brain efficiency hypothesis (Haier, 1993; Haier et al., 1988; see Neubauer & Fink, 2009). Neubauer and Fink (2009) remarked that this hypothesis is connected with the processing resources required for effectively completing a cognitive task. Our finding suggests that top performers did have the required processing resources, and, therefore, their CT follows standard developmental changes, whereas standard performers did not. In the latter case, the sustained effortful processing across sessions might be behind the observed thickening. Note that Tang et al. (2010) found negative correlations between intelligence and brain activation when completing the n-back task: the higher the intelligence, the smaller the activation.

The comparison between standard performers and controls showed statistically significant results in ventral frontal, dorsolateral prefrontal, middle, and medial temporal ROIs (Figure 7.14), although only differences in ventral frontal and middle temporal survived the Bonferroni correction. Parietal regions did not show any appreciable difference between these groups. Regions where standard performers showed preservation/thickening have been associated with working memory related processes. In this regard, the meta-analysis by Duncan and Owen (2000) revealed that regions in the frontal lobe relevant to working memory are the dorsolateral prefrontal cortex, inferior ventral frontal cortex, and dorsal anterior cingulate cortex. Specifically, the right ventral frontal cortex has been related to the inhibition of responses required to cancel an intended movement (Aron, Robbins & Poldrack, 2004), while it has been suggested that the right dorsolateral prefrontal cortex connects working memory with complex reasoning (Prabhakaran et al., 2011). There are studies suggesting that the right temporal cortex acts as an interface between the dorsal and ventral streams for visual processing, allowing the exploration of both object-related and space-related information (Karnath, 2001). Moreover, as is the case with the right inferior ventral cortex, the middle temporal lobe is relevant to interference resolution (Kirwan, & Stark, 2007; Yassa, Mattfeld, Stark, & Stark, 2011; Yassa & Stark, 2011), particularly taxed during n-back training.

7.4.3. *Changes in Cortical surface area.*

Cortical surface area changes were weaker than cortical thickness variations (Figure 7.13), which is partially consistent with developmental studies (Østby et al., 2009; Raznahan et al., 2011). The control group showed small expansions in right dorsomedial frontal and anteromedial temporal cortices, along with small contractions in right temporal (superior and middle), right pars opercularis, left precuneus, and superior temporal cortex. It has been suggested that changes in cortical surface area have a cubic relationship with age (Wierenga et al., 2014), showing contractions in 18-20 year olds. Zhou et al., (2015) reported generalized expansions (71%) in one period ranging from 10 to 20 years of age. However, CSA changes might show distinguishable trajectories by brain region. In this respect, Burgaleta et al. (2014) found CSA contractions were associated with age in right occipital and temporal regions, while surface expansions were found in bilateral dorsomedial prefrontal cortex, left ventromedial frontal, and left motor cortex. Our results for the control group are relatively consistent with Burgaleta et al., (2014), since we found CSA expansions in right dorsomedial frontal cortex and CSA contractions in the right temporal area. CSA changes were absent in the training group, except for a small expansion observed in the right middle temporal cortex. Top performers showed expansions in left posterolateral temporal cortex, whereas standard performers showed expansions in this same region of the right hemisphere. Standard performers showed expansions in left medial temporal cortex and right orbitofrontal cortex, while contractions in left middle temporal cortex were found in top performers.

CSA developmental changes may be related with changes in gyrification (Wierenga et al., 2014). Nevertheless, developmental studies addressing gyrification changes show disparate results: a) increased folding with age (Blanton et al., 2001), and b) neurogenesis and neural migration are almost complete after gestation, and, therefore, sulcal architecture is completed at birth (Hill et al., 2010). Changes in cortical surface area may be attributed to mechanistic pressures exerted by the size and complexity of the dendritic arbors (Hill et al., 2010), the size of intracortical elements, or the volume of white matter adjacent to a given gyrus or sulcus (Feczko, Augustinack, Fischl, & Dickerson, 2009).

Group differences in CSA changes were found in right pars opercularis (top training > control) and right posterolateral temporal cortex (standard training > control) and left medial temporal cortex (standard training > top training) with a moderate effect size. The posterolateral temporal region can be considered as belonging to the middle temporal and to the inferior parietal regions. As noted above, the middle temporal lobe supports working memory related processes (e.g. Hampson, et al., 2006; Zou et al., 2013), and interference resolution (Kirwan, & Stark, 2007; Yassa et al., 2011; Yassa & Stark, 2011). Moreover, the inferior parietal cortex is considered one of the most important regions supporting intelligence and high-level processing (Jung & Haier, 2007; Vendetti, & Bunge, 2014). The pars opercularis have been associated with inhibition process in go-no go tasks (Forstmann, van den Wildenberg, & Ridderinkhof, 2008). Finally, the medial temporal (fusiform region) is considered as one of the main regions of the first stage comprised by the P-FIT model (Jung & Haier, 2007).

The standard training group showed expansions (when compared to controls) in regions related with intelligence/working memory and interference resolution. The top training group showed expansions in regions supporting inhibition, while the difference between top and standard performers were found in regions related with the processing of sensory information. These latter findings support the view that only standard performers required sustained effortful processing across training sessions.

In summary, regions where the training group showed thickening or preservation, along with CSA expansion or preservation, with respect to controls, were located in right ventral frontal cortex (CT), right pars opercularis (CSA), right middle temporal cortex (CT and CSA), and one small region in the right inferior parietal region. Standard performers drove these changes. The highlighted regions are known to support working memory, interference resolution, inhibition, and reasoning. These abilities and skills are thought to contribute to n-back performance (Jaeggi et al., 2010b). Changes in the right middle temporal lobe were observed for CT and CSA, which suggest that the training regime did impact mainly on the temporal lobe (Hsu, Buschkuehl, Jonides, & Jaeggi, 2013). Takeuchi et al. (2011) found gray matter volume reductions (in bilateral dorsolateral prefrontal cortex, right inferior parietal lobule, left paracentral lobule, and left superior temporal gyrus) after adaptive working memory training based on mental calculations, which departs from our results here. However, straightforward

comparisons with our study are not warranted because of key differences in the completed task (mental calculation vs. n-back), time of training (5 days vs. 3 months), and neuroimaging analyses (VBM vs. SBM).

CHAPTER 8: Structural connectivity and n-back training (Q2.2).

8.1. Introduction

The main topic of this section is the study of structural changes after the n-back training program, as in chapter 7. However, here we present the results for structural connectivity (SC). Structural connectivity quantifies anatomical connections between different gray matter structures through white matter tracts. Diffusion weighted images (DWI) were used to reconstruct fiber trajectories. We relied on probabilistic tractography to fulfill this research goal (Iturria-Medina et al., 2007). Afterwards, a graph theory approach was used to study the influence of the training program on several measures quantifying brain network properties.

The algorithm developed by Iturria-Medina and colleagues is explained in the next section. To our knowledge, there are no studies addressing SC changes after cognitive training. Graph-based analyses were performed at two levels, nodal and global. Moreover, the network-based statistic (NBS) package (Zalesky et al., 2010b) was used to study whole-brain structural organization separating sets of highly interconnected regions (sub-networks) instead of paired regions.

These were our hypotheses: **(a) Node level.** Frontal and parietal regions will be the brain regions more sensitive to the training program, since previous research has found significant changes in white matter in these regions (Takeuchi et al., 2010; Lövdén et al., 2010). **(b) Global level.** Increased brain network communication as denoted by improved small-world properties (Sporns & Zwi, 2004). **(c) NBS.** Increases in connectivity in specific network related to the training program. Yeo et al. (2014) found that n-back performance is related to bilateral prefrontal, bilateral parietal and left middle temporal regions.

Again, only images passing quality control were analyzed here (N = 50).

8.2. Network reconstruction.

Figure 8.1 shows the main steps for network reconstruction.

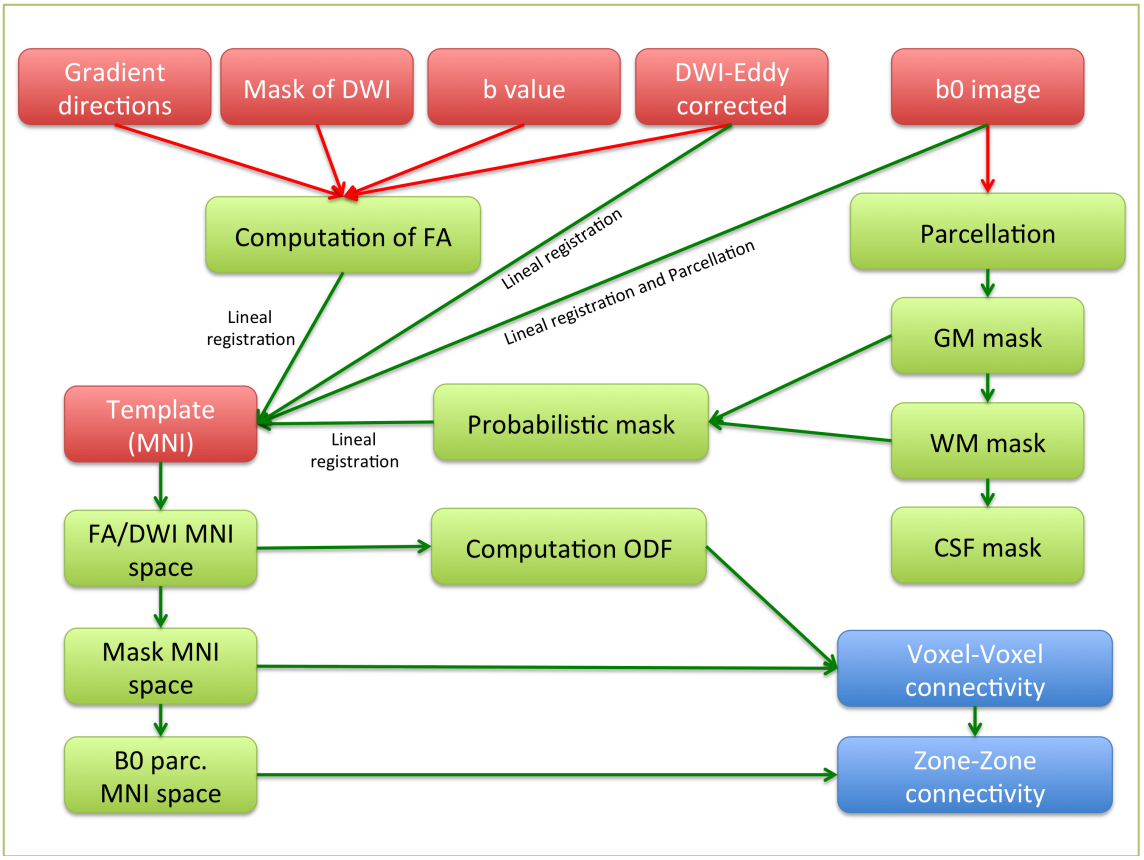


Figure 8.1. Main steps for network reconstruction. DWI = diffusion weighted image; FA = fractional anisotropy; GM = gray matter, WM = white matter, CSF = cerebrospinal fluid. ODF = orientational distribution function. MNI = Montreal Neurological Institute. The inputs are represented in red, the computed images in green and the outputs in blue. Big rectangles depict a computation process and small rectangles depict inputs or outputs of the process.

(1) **Computation of fractional anisotropy (FA).** FA was computed using the following formula:

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$$

FA can be considered an indicator of eccentricity of the ellipsoid. When the ellipsoid is long and narrow, the movement of water molecules is highly anisotropic (takes place along a preferential direction). Spherical ellipsoids reflect an isotropic diffusion (circular movement). FA can take values between 0 and 1; values close to 1 indicate high anisotropy, whereas values close to 0 express low anisotropy or isotropy. When these values are coded into a gray scale (0 = black, 1 = white), it generates an image in which areas of high anisotropy (fibers or white matter tracts) stand out from a dark

background (gray matter, cerebrospinal fluid). To further distinguish the distribution of fibers' direction, it is common practice to encode the direction of principal eigenvector using a standard color map. Color represents fiber orientation: in the dorsal-ventral (top-bottom) blue fibers appear, green is plotted in the anterior-posterior axis, and red represents fibers from left to right or vice versa.

FA computation was carried out using TBSS (Tract-Based Spatial Statistics) (Smith et al., 2006) within the FSL program (Smith et al., 2004). The required inputs were: DWI image, b value, gradient directions files and the mask generated during preprocessing (see Figure 8.2).

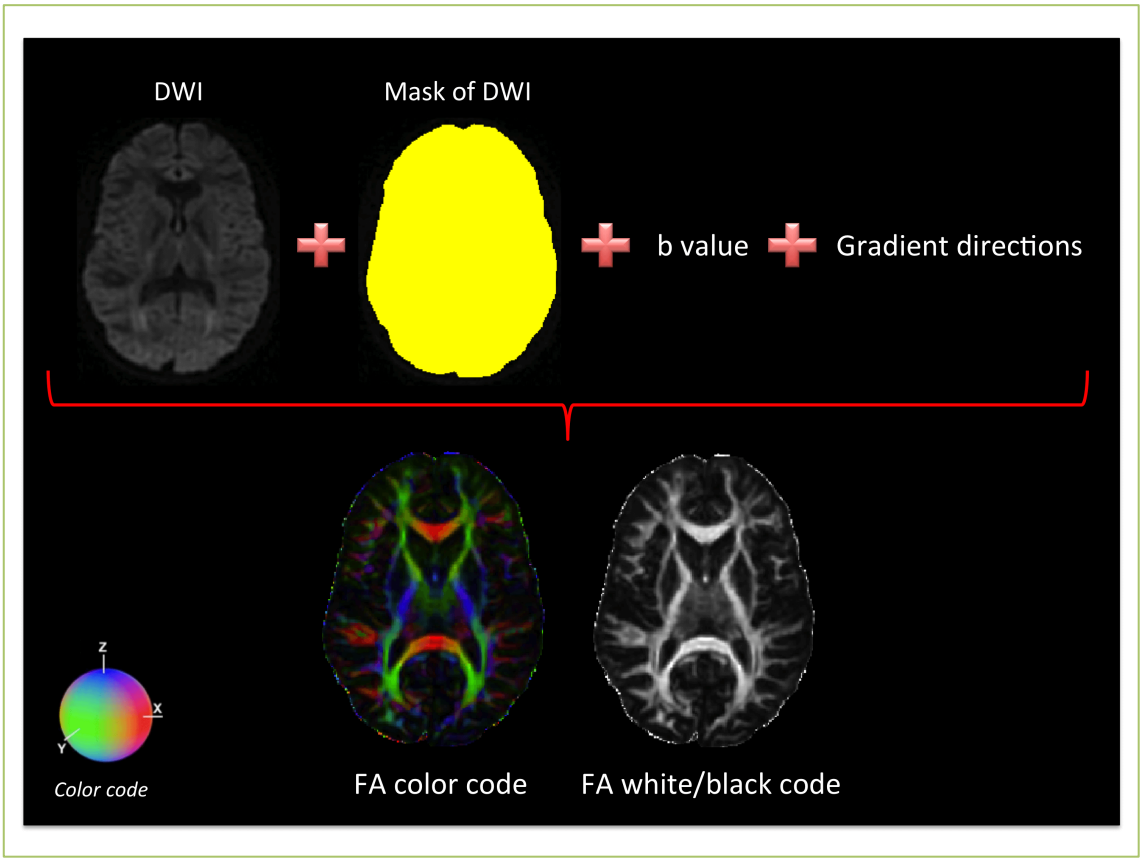


Figure 8.2. Example of FA computation with FSL (Smith et al., 2004). DWI = diffusion weighted image. In FA with white/black code, white represents regions with high anisotropy (white matter) and black represents low anisotropy (gray matter and cerebrospinal fluid). In FA with color code, blue represents top-bottom fibers (and vice versa), red represents right-left fibers (and vice versa) and green represents anterior-posterior fibers (and vice versa).

(2) Image **b0 was segmented in 3 tissue probability maps**: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF), using SPM8 software package (available at

http://www.fil.ion.ucl.ac.uk/spm/). Afterwards, a function developed by Iturria-Medina and Valdés-Hernández (2005) was used to calculate a three dimensional mask that only possesses nonzero values for white matter and gray matter points (see Figure 8.3).

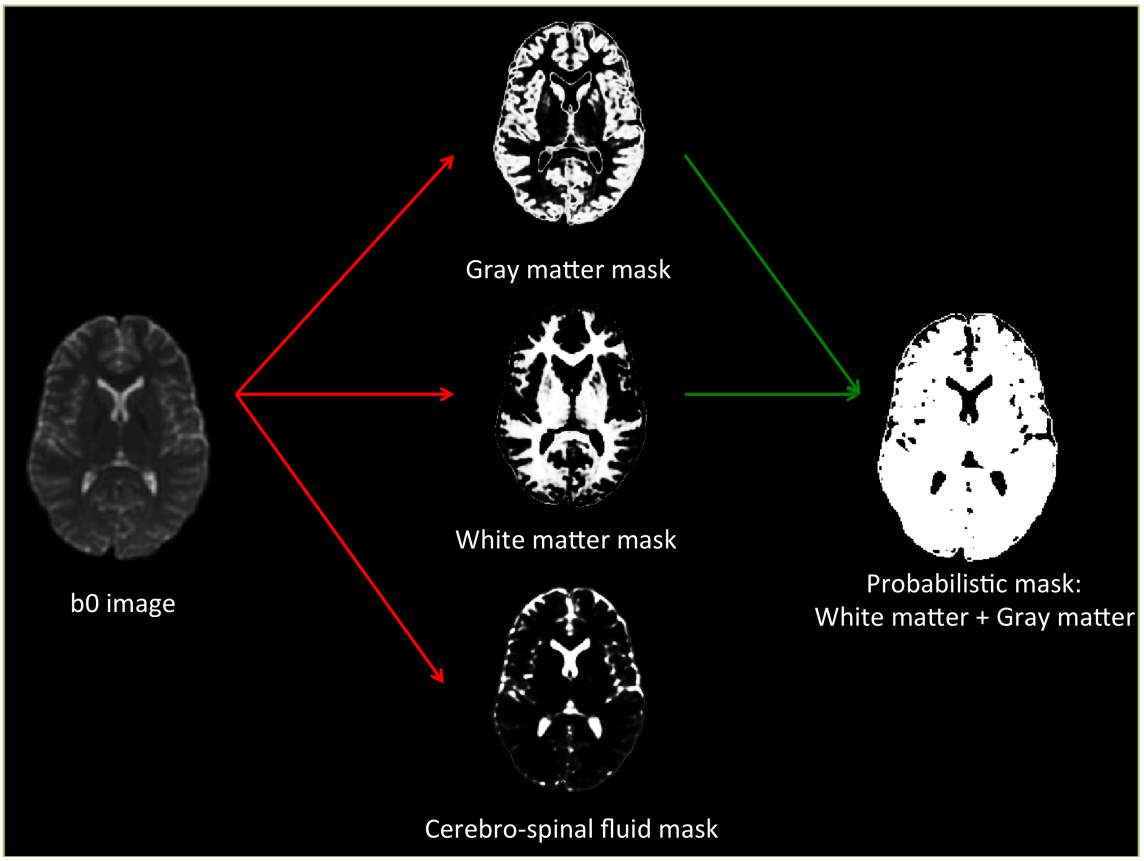


Figure 8.3. Example of segmentation in white matter, gray matter and cerebrospinal fluid. After, masks of gray matter and white matter were combined to generate a three dimensional mask.

(3) The third stage involved **registration of probabilistic mask, FA, DWI and b0 images to template space** (MNI space) applying a linear transformation. Also, each b0 image was automatically parcellated into 62 GM regions using the OASIS-DKT 31 template (Klein & Tourville, 2012) and *Freesurfer* (Fischl, 2012) segmentation of subcortical structures and cerebellum (Fischl et al., 2002). These divisions were employed to define the template (see Figure 8.4). Therefore, each brain was parcellated into 83 regions of interest (62 corticals + 14 subcortical + 5 Cerebellum + 2 basal forebrain).

(4) Finally, **a graph-based tractography algorithm** (Iturria-Medina et al., 2007) was used. This algorithm assumes that the brain is a graph (Hagmann et al., 2006). A graph

G=[**N, A**] is defined by a set **N** of **n** elements called nodes and a set **A** of elements called arcs/links or edges. Each arc **a**∈**A** is assigned a number **w(a)**∈**R** (belonging to real numbers), indexing the weight of the arc, where each arc weight is considered as the probability of its existence. In the brain, these concepts are translated into:

$\mathbf{G_{brain}}$ = [**N, A**] where **A** is the set of white matter links between contiguous voxels in **N**; **N** = set of voxels having a non-zero probability of belonging to some cerebral tissue. Then, $\mathbf{G_{brain}}$ is a discrete set of points (nodes) representing voxels and a set of lines (arcs or edge) representing connections between contiguous voxels and the **weight (w)** of an arc is the probability that linked nodes are really connected by nervous fibers (Iturria-Medina et al., 2007).

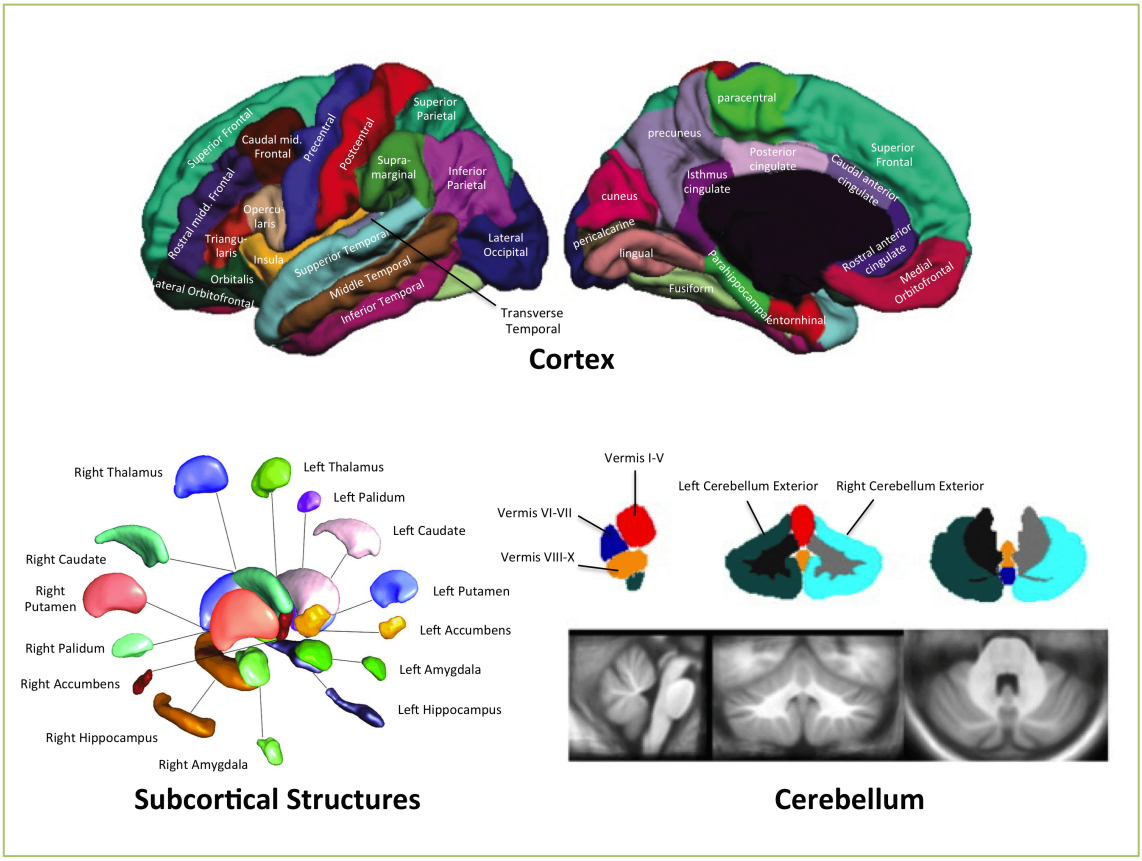


Figure 8.4. Top panel shows the gray matter parcellation of OASIS-DKT 31 template (Klein & Tourville, 2012). Bottom panel depicts the Freesurfer parcellation of subcortical structures (right) and cerebellum (left) (Fischl et al., 2002).

The steps to compute the weight of each arc were:

First, combination of the information of probabilistic tissue segmentation of the brain and reconstruction of the q-ball orientation distribution function (ODF) obtained of

DWI image in MNI space (Tuch, 2004). This ensures that only those pairs of nodes with high probability of belonging to gray/white matter and high probability of sharing fibers will have higher weights, which is equivalent to having high probability of being connected. From the ODF, we computed generalized fractional anisotropy (GFA) (Tuch, 2004), which is a HARDI anisotropy measure similar to the popular DTI fractional anisotropy (FA). HARDI models allow the estimation of more than one main direction at each voxel and therefore it is better suited to solve the problem of crossing-fibers. GFA was used to define arc weight.

Afterwards, fiber tracking and **node-node connectivity** was computed looking for "*the most probable path problem*" between nodes of interest in the defined Brain Graph. An iterative algorithm was employed to find the most probable trajectory, which is considered to evaluate the real existence of fiber pathways between these points[1]. The anatomical connectivity measure is defined as the lowest weight of the links belonging to the most probable path (shorter path).

Third, the evaluation of connectivity between clusters of nodes (**zone-zone connectivity**) in $\mathbf{G_{brain}}$ was used to define connectivity measures between brain anatomical areas. The anatomical areas were defined with the cortical parcellation (OASIS-DKT 31), subcortical and cerebellum regions. In this context, the edges are defined directly between nodes of different areas. Three different anatomical connectivity measures were computed with this algorithm: (a) **Anatomical Connection Strength** (ACS) gives an estimate of the amount of nervous fibers shared by these areas, (b) **Anatomical Connection Density** (ACD) is searched as a measure of the fraction of the surface involved in the connection with respect to the total surface of both areas. Thus, it can be estimated as the ACS relative to the number of nodes belonging to the surfaces involved in the computation, and (c) **Anatomical Connection Probability** (ACP) is searched as a measure of the probability of two areas to be connected at least by a single connection.

An example of output of node-node and zone-zone connectivity is represented in Figure 8.5. A node-node connectivity 3D-image was computed for each region defined by the employed template. In these maps, the probability of connection between voxels within

[1] In order to consider physiological and anatomical evidences of fiber bundles a function, which penalizes path curvature, shape in the brain is used as priori information.

regions and the rest are represented in a color scale (see Figure 8.5). Zone-zone connectivity is a symmetric matrix (*N*x*N*) where each node is represented in a row and column; again, a color code is used to represent the probability of connection between regions (see Figure 8.5). Finally, three indicators were obtained (ACD, ACS and ACP) but only ACD-matrices were analyzed since there is a high correlation between indices in healthy subjects (Iturria-Medina et al., 2007) and ACD-matrices or "*normalized matrices*" are the most used.

In the next section, the analysis computed to answer the main question ("*Were there changes in structural connectivity after n-back training program?*"), are detailed.

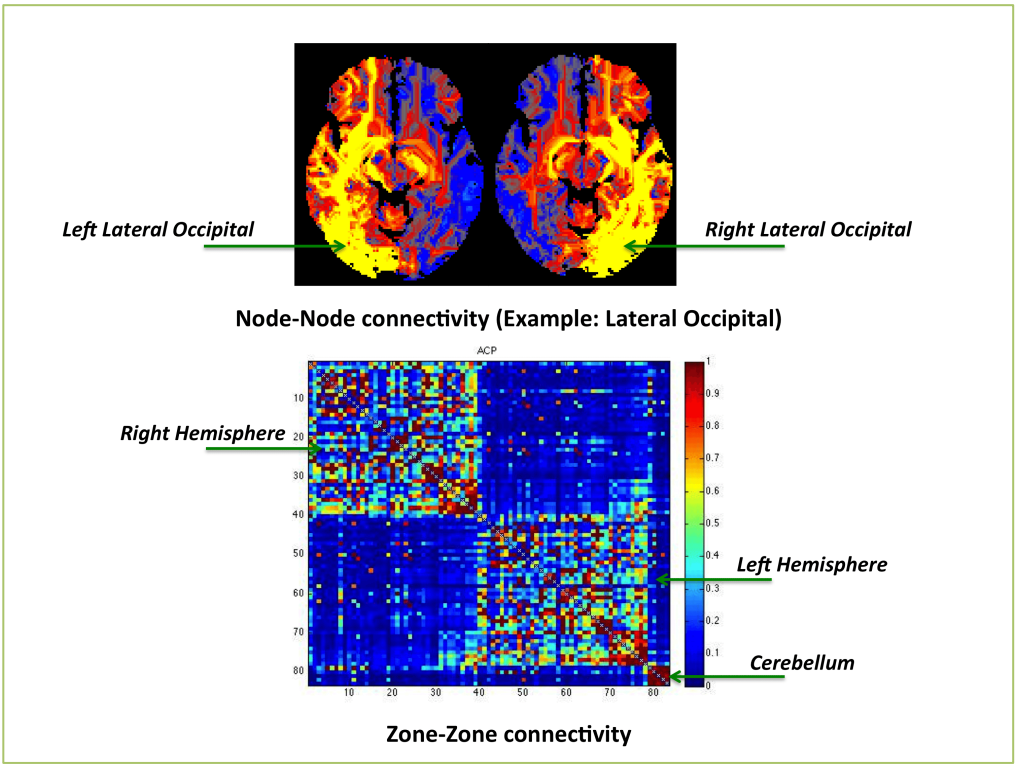


Figure 8.5. Top panel shows an example of node-node connectivity (probability maps: yellow = high probability of connection; blue = low probability of connection). Bottom panel displays the zone-zone connectivity matrix. Probability of connection between each region of interest (ROI) is represented with a color code from blue (low probability of connection) to red (high probability of connection).

8.3. Changes in connectivity at the node level.

First, descriptive statistics (mean and standard deviation) of ACD matrix were computed for each subject at the node level before and after training. Generalized fractional anisotropy (GFA) was used to define arc weight. Information of zone-zone

matrix per subject was analyzed to compute these descriptive statistics. The connectivity of each node was represented with a numerical value (from 0 to 1). Thus, the mean connectivity between each node with the rest was computed (the connection of the node with itself was not shown).

Figures 8.6 and 8.7 show the descriptive statistics: mean and standard deviation respectively of 15 top nodes with the highest connectivity value. The nodes with high connectivity were similar before training in both groups. For the training group, these hubs were: bilateral superior frontal, superior parietal, precuneus, thalamus, putamen and pallidum, right caudate, hippocampus, and paracentral. For the control group, the hubs were: bilateral superior frontal, superior parietal, precuneus, thalamus, putamen, pallidum, paracentral, and left hippocampus. These results are consistent with previous research analyzing structural brain network (van den Heuvel & Sporns, 2011, 2013).

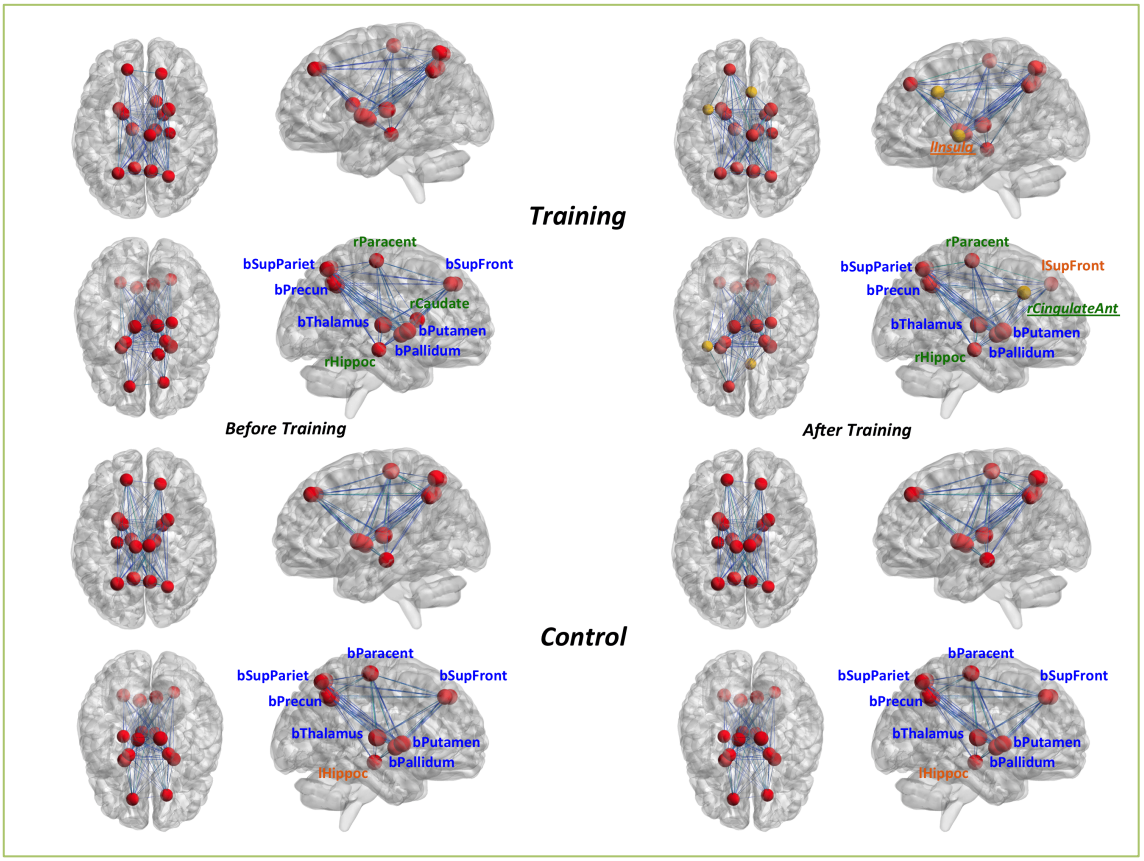


Figure 8.6. Mean in connectivity strength of 15 top regions (18%) in the brain network for the group of participants (Top Panel –training group, Bottom Panel –control group). Golden globes depict the new hubs after training. Labels in blue depict bilateral nodes, right nodes are represented in greenand the left nodes are represented in orange.

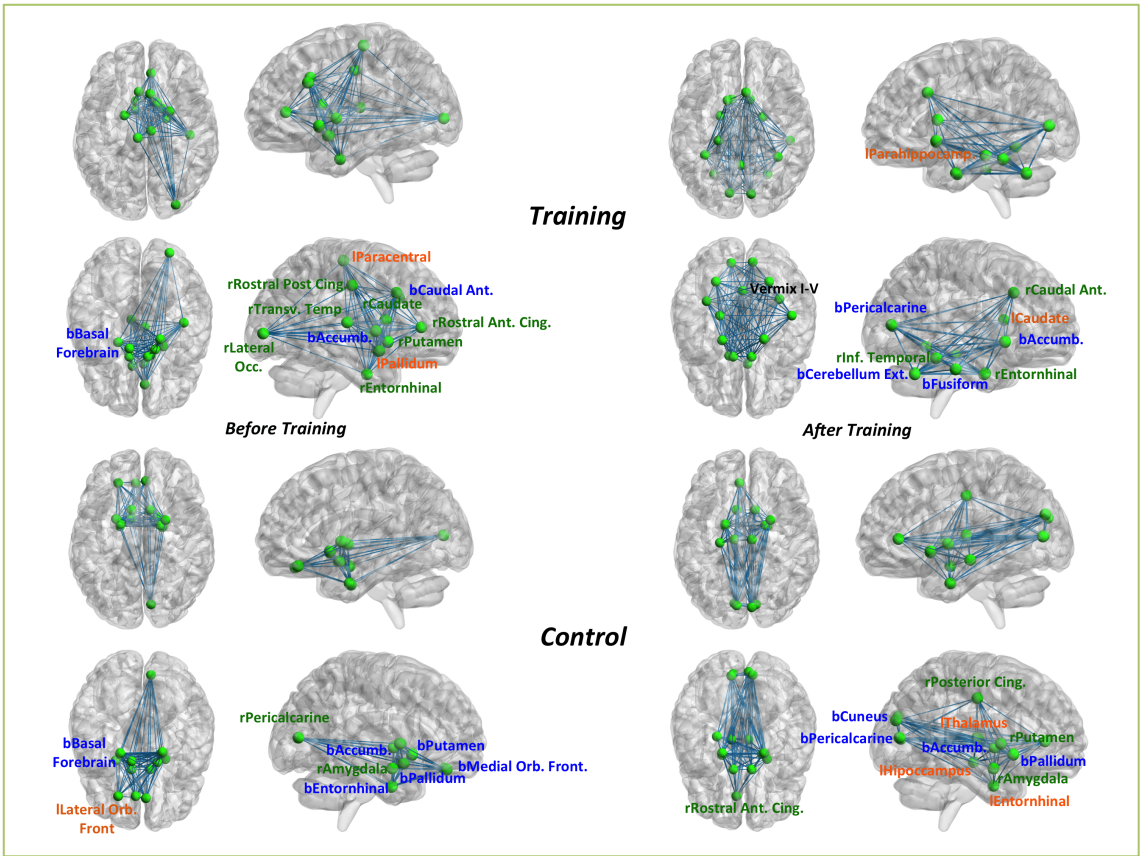


Figure 8.7. Standard Deviation in connectivity strength of 15 top regions (18%) in the brain network for the group of participants. (Top Panel –training group, Bottom Panel – control group). Golden globes depict the new hubs after training. Labels in blue depict bilateral nodes, the right nodes are represented in green , left nodes in orange and cerebellum nodes in black.

Top nodes remains unaltered in the control group from the first to the second MRI scan, whereas some descriptive changes were observed in the training group. Specifically, the left insula and the right anterior cingulate increased their connectivity level after training. It is noteworthy that standard deviation values were less stable than mean connectivity values.

Secondly, the correlations between n-back performance and connectivity strength at the node level were computed. No statistically significant results were found for the visual (mean session 1 to 4), auditory (mean session 1 to 4), and dual session (block 1: mean session 1 to 4; block 2: mean session 5 to 8: block 3: mean session 9 to 12 and block 4: mean session 13 to 16).

Thirdly, standardized change scores were computed (Jaeggi et al., 2011) [(after training – before training) / SD (before training)] for each connection within each individual

ACD matrix, using GFA to define arc weight. The outcome was a matrix where each number represented the standardized changes of the connection between $node_x$ and $node_y$ ("x" and "y" are real numbers from 1 to 83 – regions of interest of template employed). Connectivity between voxels of the same node was not computed (the main diagonal was comprised by zero values). Afterwards, the mean per column was computed to estimate the standardized change for each node and for each individual. The descriptive results (mean) per group (training vs. control) are shown in Figure 8.8.

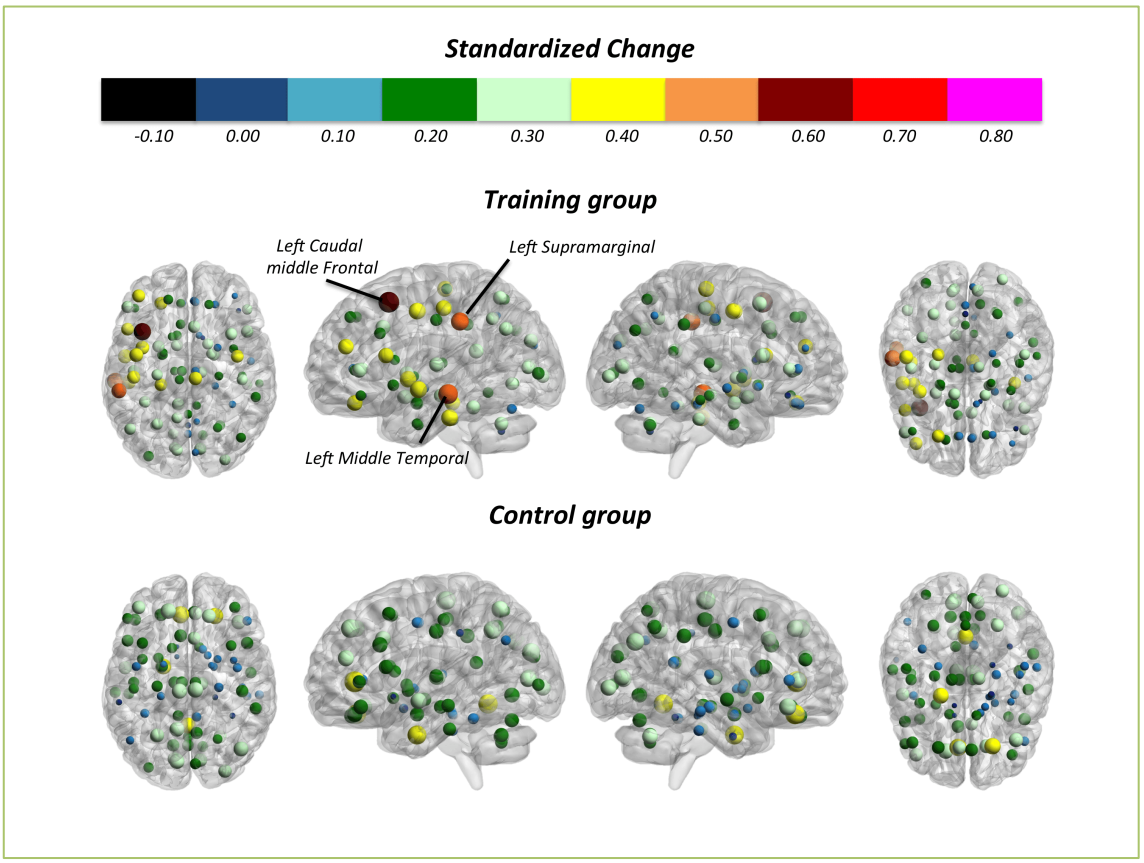


Figure 8.8. Standardized Change [(after − before)/ SD at the before] for training group and control group for all nodes. Size and color codes represent the magnitude of standardized change.

Standardized changes were smaller for controls, especially in anterior nodes of the left hemisphere. The training group showed moderate changes ≥ 0.40 (yellow nodes) in several relevant nodes: left caudal middle frontal, left supramarginal, and left middle temporal nodes.

Afterwards, ANCOVA analyses were computed to examine whether the standardized change showed statistically significant group differences. The dependent variable was

the standardized change, group was the independent variable, and the covariate was the connectivity of each specific node before training. Thus, the model was computed 83 times (one per node of template employed). Results are shown in Figure 8.9.

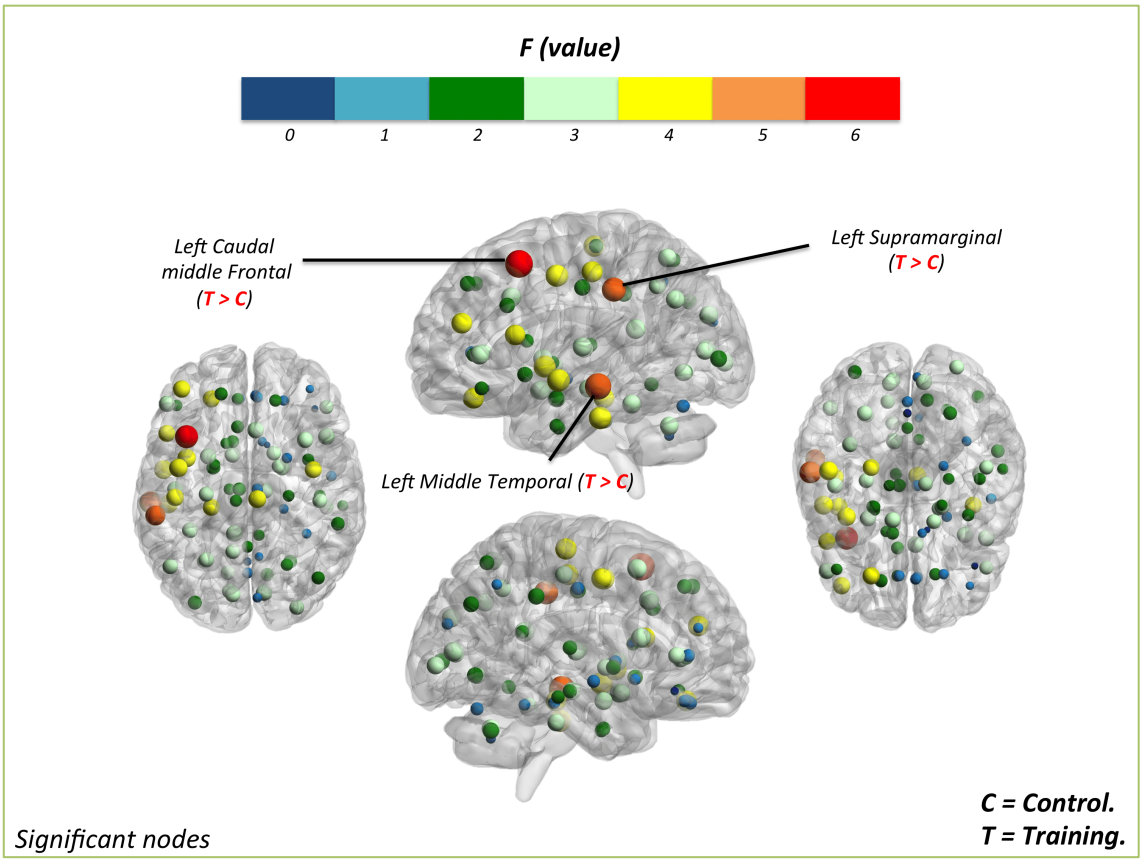


Figure 8.9. ANCOVA results from the comparisons of standardized change between training and control group. Dependent variable was the standardized change, independent variable was the group and covariate was the connectivity of the specific node before n-back training program.

The left caudal middle frontal, left supramarginal, and left middle temporal nodes showed group differences, although results failed to pass Bonferroni corrections.

Finally, comparisons between top performers, standard performers, and control groups were computed. Figure 8.10 shows the mean standardized changes for each group at each node.

Results for top performers and controls were alike, whereas standard performers showed increments in connectivity at the left hemisphere: left caudal middle frontal (standard change = 0.80), left inferior and left middle temporal (standard change = 0.60). The right paracentral also showed a high standard change = 0.60. The results

from the ANCOVA testing the statistical significance of group differences are shown in Figure 8.11.

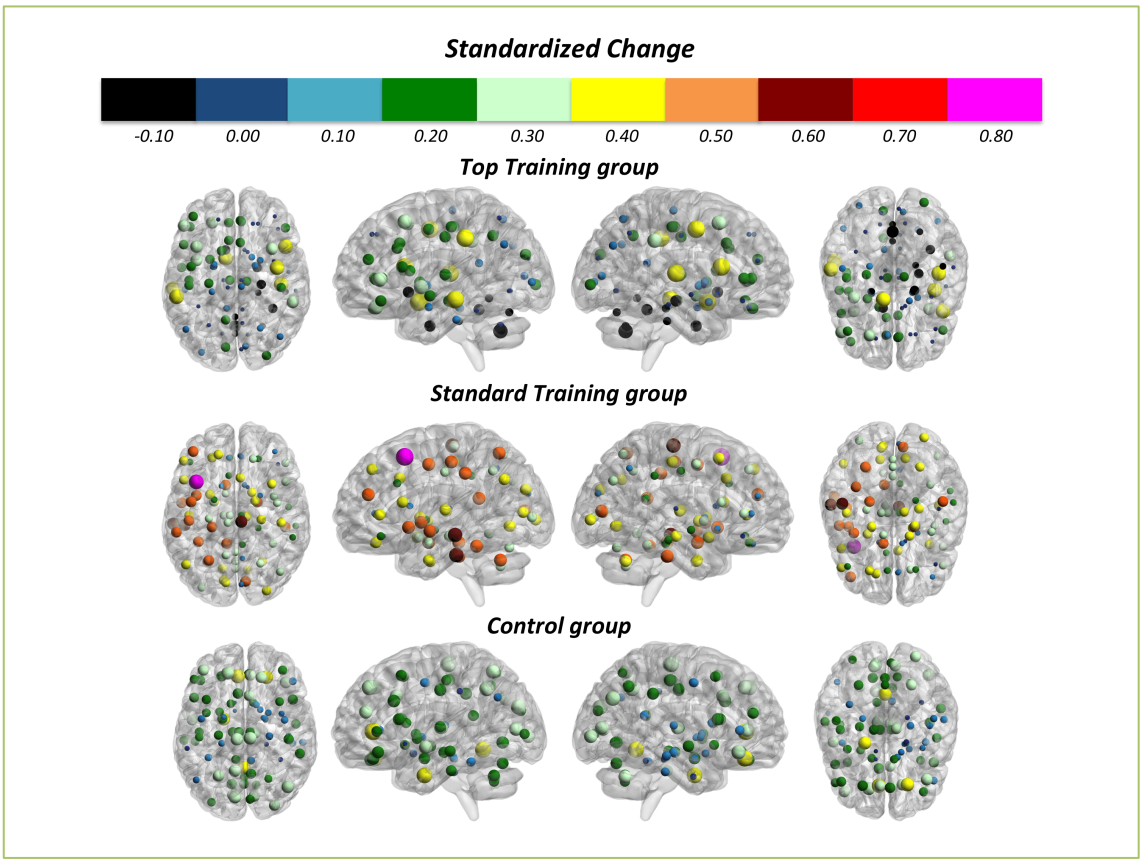


Figure 8.10. Standardized Change [(after − before)/ SD at the before] for top training group (top), standard training group (middle) and control group (bottom) for all nodes. Size and color codes represent the magnitude of standardized change.

Again, left middle temporal, supramarginal, and caudal middle frontal regions revealed greater changes in the training group, but this difference was driven by standard performers. Post-hoc analyses showed a failure to pass the Bonferroni correction, and, therefore, caution is required.

In conclusion, standard performers showed generalized connectivity increments in the left hemisphere (standardized change ≥ 0.40). Left supramarginal, left caudal middle frontal, and left middle temporal nodes are highlighted, although findings failed to survive to correction for multiple comparisons.

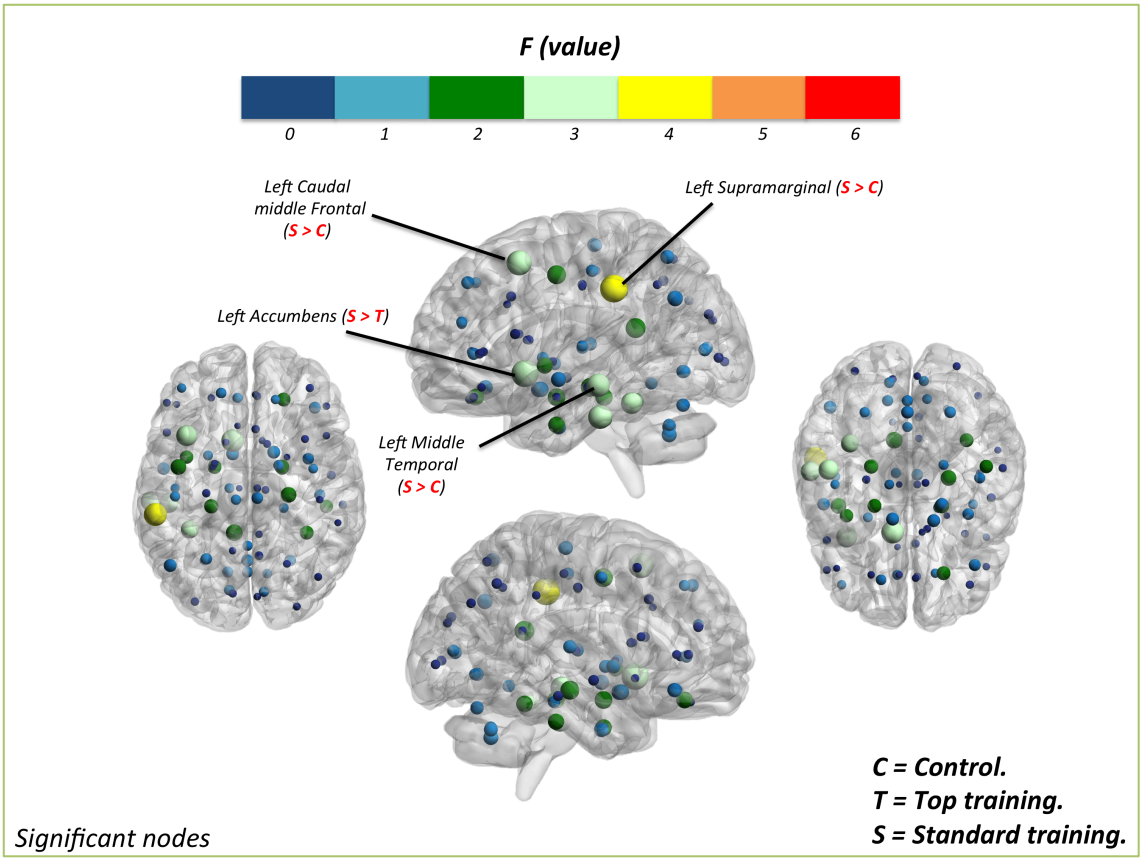


Figure 8.11. The results from the ANCOVA comparing the standardized change in top training, standard training and control group. The dependent variable was the standardized change, the independent variable was the group and the covariate was the connectivity of the specific node before n-back training program.

8.4. Changes in connectivity at the global level: small-world properties improvements.

We chose to analyze characteristic path length (L) and mean clustering (Cl). These indices allow the quantification of the "small-worldness" of a network (Sporns & Zwi, 2004). Small worldness implicates a network highly connected (⇑Cl) with short average steps (⇓L) connecting different sub-networks.

We computed Pearson correlations between these two indices at baseline (before training) and training performance. Again, no statistically significant correlations were found. Specifically, the range of correlations between Cl and training performance ranged from -.20 (visual) to .06 (dual sessions 13-16), and the range of correlation for L was from -.09 (dual sessions 1-4) to .18 (visual).

Descriptive results for Cl and L in the training (whole group, top performers, and standard performers) and control groups before and after training are shown in Table 8.1. Also, the mean standardized change for each measure per group are included.

Table 8.1. Descriptive statistics for clustering (Cl) and characteristic path length (L) in the training and control groups before (baseline) and after training (posttest). Standardized Change [(After – Before)/SD(before)] is also reported. Stnd = Standard. M = mean, SD = standard deviation.

	Cl: *M(SD)*		Std. Change	L: *M(SD)*		Std. Change
	Before	**After**		**Before**	**After**	
Training	.020(.003)	.022(.002)	0.59	47.82(6.9)	43.49(4.2)	-0.52
Control	.020(.004)	.021(.005)	0.47	47.52(9.6)	46.18(12)	-0.16
Top Training	.020(.003)	.021(.003)	0.32	47.01(4.6)	44.51(5.1)	-0.30
Stnd. Training	.020(.003)	.022(.002)	0.76	48.33(8.1)	42.85(3.4)	-0.66

The training group showed greater Cl increments than controls, along with greater L decrement. Again, the standard performers drove this difference. ANCOVA analyses were computed to test for statistical significance. Regarding the comparison training vs. controls, Cl increment was not statistically significant, $F(1,48) = .203$, $p = .327$, $\eta_p^2 = .004$. Similar results were found for L, $F(1,48) = 1.18$, $p = .141$, $\eta_p^2 = .024$. This remained the same with respect to the comparison among top performers, standard performers, and controls (Cl, $F(2,47) = .350$, $p = .353$, $\eta_p^2 = .015$ and L, $F(2,47) = .702$, $p = .501$, $\eta_p^2 = .029$).

Therefore, no statistically significant changes were observed at this global level, so we moved to the final analytic stage, namely, network-based analyses.

8.5. Network-based statistics (NBS).

NBS is intended to control family-wise error rate (FEW) when univariate testing is computed at every connection comprising the graph (Zalesky et al., 2010b). NBS specifies information about whole-brain structural organization separating sets of highly interconnected regions (sub-networks) instead of paired regions. This approach has been applied to compare healthy individuals and patients (Bai et al., 2012; Verstraete et al., 2011; Zhang et al., 2011). It has only recently been applied to compare groups of healthy participants (early bilinguals and monolinguals) (García-Pentón et al., 2014).

Here, we applied NBS to compare training and control groups regarding their change after the training program. Thus, two sample t-tests were computed, using the individual matrices, including the standardized changes [(after − before)/ SD at the before] for in each connection. NBS computes a two-sample t-test for each independent connection to test if the connectivity values of two groups belong to distributions with equal means (null hypothesis). Afterwards, a preselected T-value (T-threshold = 3.7) was used to limit the statistical value calculated at each link of the standardized differences matrix to identify the set of supra-threshold connections. Afterwards, interconnected components presented in the set of supra-threshold connections were identified and their size is saved. Size is the number of connections (arcs) in the set of selected connections. Therefore, connections where the null hypothesis was rejected give rise to a component. Nonparametric permutation testing (K = 5000 permutations) was used to estimate the significance of each component identified. For each permutation, the group to which each subject belongs was randomly exchanged, and then the statistical test was recalculated. Later, the same threshold was applied to create the set of supra-threshold links for each K permutation. The size of the largest component in the set of supra-threshold links derived from each K permutation was stored, providing an empirical estimation of the null distribution of the maximal component size. Finally, a corrected *p*-value of each observed connected component was computed, calculating the proportion of the 5000 permutations for which the largest component size was greater than the observed connected component size and then normalized by K. This allowed to control for the family-wise error (FWE) associated to each connected component, based on its size.

Following the steps described above, one component (sub-network) was identified. For this sub-network, standardized changes were higher in the training group than in controls (see Figure 8.12).

This sub-network included temporal, frontal, parietal, and subcortical regions, along with the insula. The most highly connected node in this network was located in the left middle temporal region, and it was highly interconnected with the bilateral basal forebrain, left parahippocampal, pallidum, supramarginal and inferior parietal, right insula, accumbens, post central, pars opercularis, and pars triangularis.

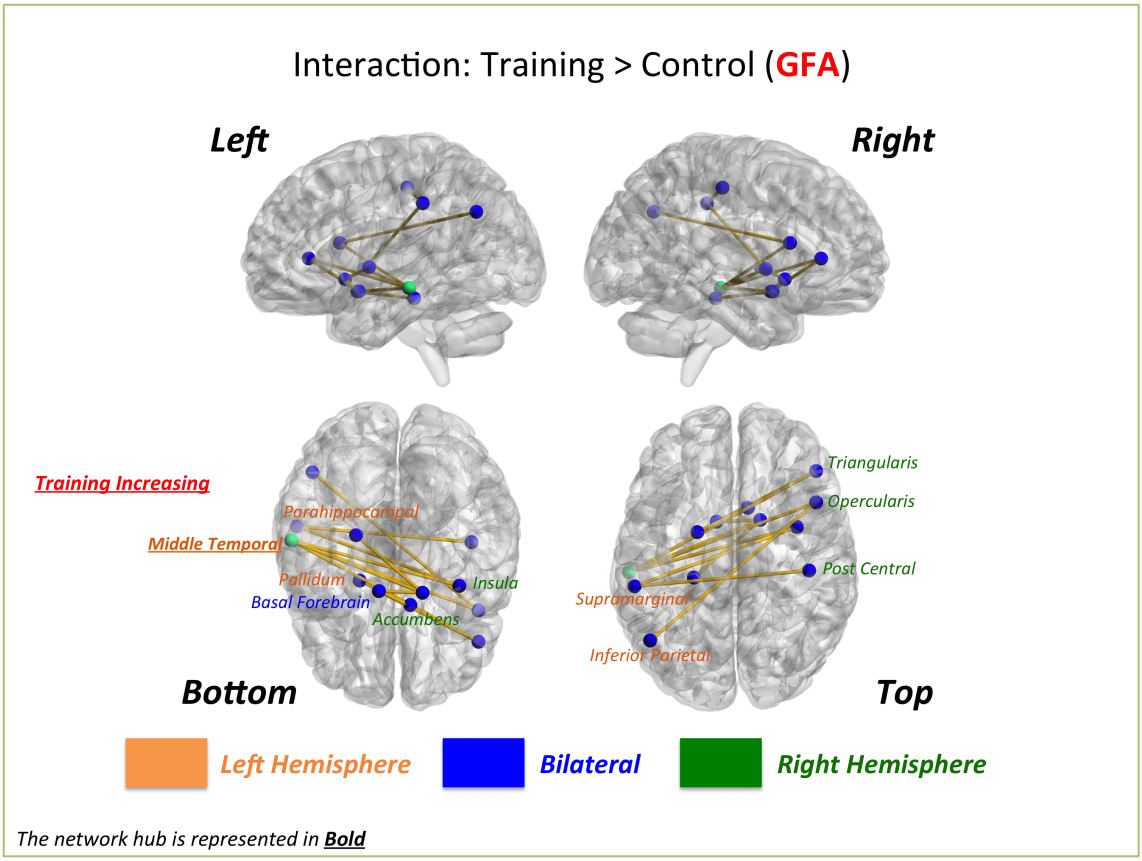


Figure 8.12. Sub-network where standardized changes were higher in the training group than in controls. Generalized fractional anisotropy (GFA) was used to define the weight of the connections.

Nevertheless, the identification of this network might be due to (a) a decrement of the connectivity in controls along with no changes in the training group, (b) an increment of the connectivity in the training group along with no changes in controls, (c) a decrement in both groups, but greater in controls, and (d) an increment in both groups, but higher in the training group. To explore this further, paired t-tests were computed separately for each group. ACD matrices (GFA as estimation of weight) obtained before and after training were submitted to this analysis. Results revealed an increment for the training group in this network along with no changes in controls.

Finally, top performers, standard performers, and controls were compared. The general pattern of results showed a lack of differences between standard and top performers. However, both groups were different to the control group. The comparison of top performers vs. controls showed practically the same network identified previously with left middle temporal as network hub (Figure 8.13).

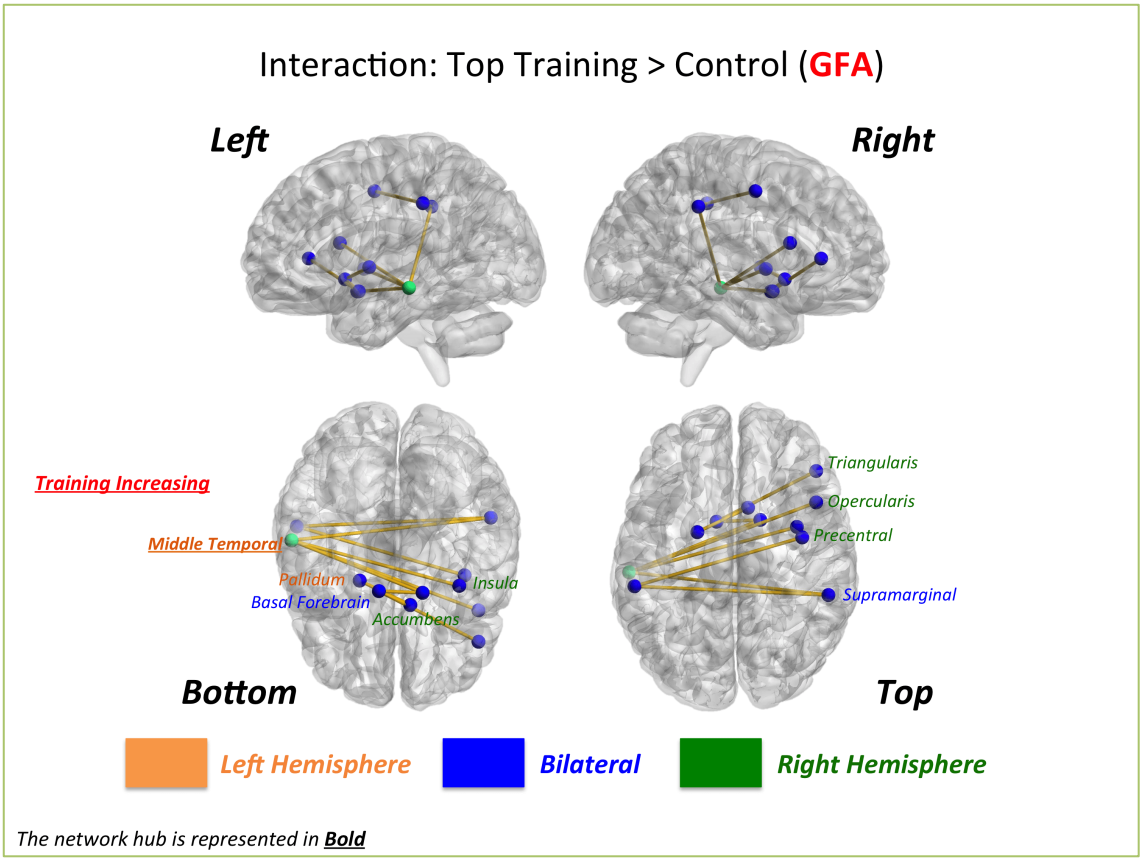


Figure 8.13. Sub-network where standardized changes were higher in top training than in control group. Generalized fractional anisotropy (GFA) was used to define the weight of the connections.

Differences between both networks were detected in several nodes: nodes were detected in the comparison between top performers and control but not in the comparison between training vs. control (right supramarginal and precentral), and vice versa (right parahippocampal, inferior parietal, and postcentral). Paired t-tests showed an increment in this sub-network for top performers.

Figure 8.14 shows the sub-network, where standard training group showed an increment higher than in the control group.

Again, the results found in this contrast were highly similar to those found in the global contrast (training vs. control). These were the discrepant nodes: right pallidum, left insula and right amygdala were detected in comparison to standard performers vs. control and not in comparison with the whole sample; and left inferior parietal, supramarginal and right opercularis and postcentral were found in comparison to whole sample and not in the comparison of the standard training group with the control. Again,

left middle temporal was the hub of the sub-network. Pair t-test analyses showed that an increment was found in this sub-network in the standard training group.

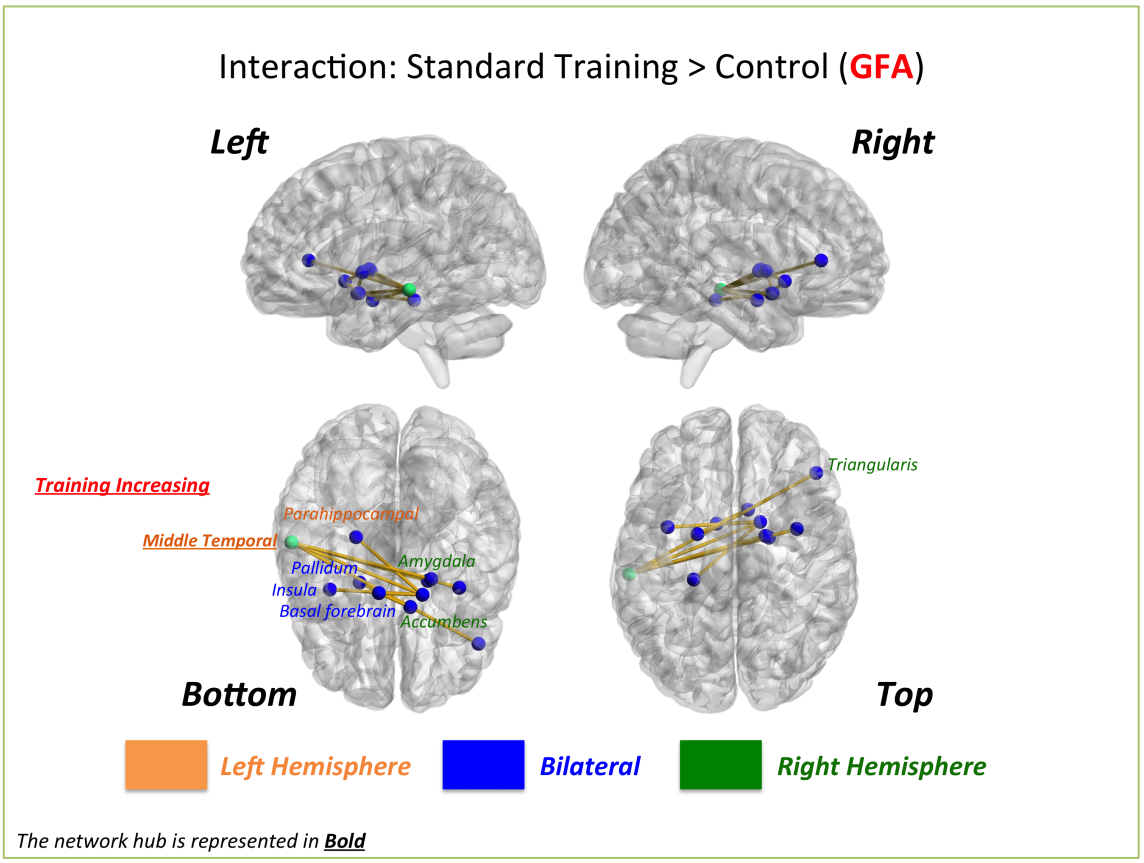


Figure 8.14. Sub-network where standardized changes were higher in standard training than in control group. Generalized fractional anisotropy (GFA) was used to define the weight of the connections.

In summary, an increment in a sub-network was found for the training group in comparison to the control group. Nodes involved in this sub-network were: bilateral forebrain, left parahippocampal, left pallidum, left supramarginal, left inferior parietal, right insula, right accumbens, right postcentral, right pars triangularis, right pars opercularis and left middle temporal. Right middle temporal was the core in this network. A 75% of this sub-network was shared with the sub-network identified in the comparison of top training vs. control group. However, right supramarginal (parietal) and right precentral (frontal) were detected as relevant in the comparison of top > control. Finally, the network detected in the 'standard training vs. control' comparison shared a 67% of nodes found in the 'training vs. control' comparison. The new nodes detected were the left insula, right pallidum (subcortical) and right amygdala (subcortical).

8.6. Discussion.

Changes in structural connectivity were analyzed using three approaches: (a) changes at the node level, (b) graph-theory indices and (c) network-based statistic (NBS). In all methods, the interaction (group x time) was analyzed to study whether the changes were different in the groups analyzed here.

8.6.1. *Changes in connectivity at the node level.*

The main finding was that controls showed smaller standardized changes than the training group. This difference was driven by standard performers, who showed connectivity increments in three nodes: left caudal middle frontal, left supramarginal (parietal), and left middle temporal. This is consistent with previous reports showing changes in white matter voxels at temporal and parietal regions (Takeuchi et al., 2010). Note, however, that these results failed to pass statistical correction for multiple comparisons.

The caudal middle frontal has been highlighted in a recent meta-analysis addressing the structural basis of intelligence (Basten et al., 2015). This region is part of a broad working memory system that includes the prefrontal cortex, anterior cingulate cortex, and hippocampus (Braver & Cohen, 2000).

The supramarginal gyrus (inferior parietal region) is key for supporting high-level cognitive processes. In this regard, Vendetti and Bunge (2014) underscored this region for working memory, attention processes, and episodic memory. Also, Jung and Haier (2007) highlighted the leading role of the arcuate fasciculus for intelligence. The changes detected at the supramarginal gyrus seem consistent with this perspective.

The middle temporal has not been particularly underscored for the P-FIT model as relevant to intelligence. However, Basten et al.'s (2015) meta-analytic structural results disagreed, suggesting a leading role for the temporal lobe. This later point of view is consistent with the findings here: the middle temporal was systematically the brain region more sensitive to training, as showed by cortical thickness, cortical surface area, and structural connectivity. As discussed in chapter 7, the middle temporal area has been nominated as supporting interference resolution (Kirwan, & Stark, 2007; Yassa et

al., 2011; Yassa & Stark, 2011), and this cognitive process is key for proper n-back performance.

Finally, standardized increments were observed in the left accumbens (subcortical) regarding the comparison between training groups (top vs. standard). The accumbens is thought to support motivation, emotion (pleasure), and learning (reward and reinforcement) (Cardinal, Parkinson, Hall, & Everitt, 2002; Goto & Grace, 2005; Kelley, Smith-Roe, & Holahan, 1997). In chapter 6, relevant differences between top and standard performers were reported in their perceived difficulty level (d = -.69). As expected, top performers perceived the training program as less difficult and less challenging than standard performers. Remember that the training sessions did provide systematic feedback. Therefore, we speculate that the observed connectivity increments in standard performers might be associated with operating a reinforcement process across training sessions.

8.6.2. *Changes in connectivity at a global and sub-network level.*

Graph-theory results provided null results. However, the training group showed larger increments in clustering (Cl) and greater decrements in characteristic path length (L), which supports improvements in small-worldness, at least as a trend.

NBS allowed the detection of a sub-network more sensitive to change presumably by the influence of training. This network comprised subcortical (bilateral forebrain, left pallidum, right accumbens), frontal (right pars triangularis and right pars opercularis), parietal (left supramarginal, left inferior parietal and right postcentral), and temporal regions (left parahippocampal, and left middle temporal), along with the right insula. These regions were highlighted for the n-back task in the meta-analysis of functional studies conducted by Yeo et al. (2014) (see Figure 3.15, Chapter 3). Interestingly, the node within this sub-network showing more enriched connectivity with the rest of the nodes was the left middle temporal region. Thus, this area might serve as an important way station for communication across the specific regions included in the sub-network identified.

The nodes of this network can be divided according to their main functions:

(a) Interference resolution. Nodes of this network are involved in inhibition processes (right pars triangularis and right pars opercularis; Aron et al., 2004), interference resolution (left middle temporal; Kirwan, & Stark, 2007; Yassa & Stark, 2011) and inhibition of movements (left pallidum; Dostrovsky et al., 2000). The most central node of this sub-network, the middle temporal region, is connected directly with other regions related with inhibition processes (right pars triangularis, right pars opercularis and left pallidum). All these components are relevant to coping with the n-back task (Jaeggi et al., 2010b). Therefore, the results suggested an increment in the connection of regions associated with the resolution of interference, along with cognitive and motor inhibition processes. This fits the behavioral results reported in chapter 6: the compatibility effect in the spatial attention task was smaller (less interference) in the training group after the training program.

(b) Working Memory-Intelligence. Regions found in the parietal region (left supramarginal, left inferior parietal and right postcentral) have been associated with working memory and intelligence (Barbey et al., 2014a; Colom et al., 2007; Jung & Haier, 2007; Vendetti & Bunge, 2014). The meta-analysis by Basten et al. (2015) showed that the inferior temporal region (parahippocampal) is also frequently found in structural studies related with the intelligence construct. Indeed, the parahippocampal region is thought to support encoding and recognition of environmental scenes (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). Finally, the insula has been negatively associated with intelligence (Basten et al., 2015). Our results indicate that the insula is directly linked to the middle temporal node, while parietal regions are associated indirectly with the middle temporal node through the right pars triangularis and right pars opercularis. Changes in these nodes might be related with the cognitive changes observed here for the visuospatial tests and tasks (chapter 6).

(c) Motivation. The accumbens and basal forebrain belong to the ventral striatum system, highly associated with motivation (Goto & Grace, 2005; Kelley, Smith-Roe, & Holahan, 1997). As noted before, the cognitive training regime included systematic feedback. It has been suggested that this sort of feedback may foster attention processes (Schwaighofer, Fischer & Bühner, 2015). Interestingly, these nodes are directly linked with the middle temporal area.

Finally, regarding the comparison between top performers vs. controls, the right supramarginal and right precentral were highlighted nodes. These parietal and frontal regions have been related with intelligence (Jung & Haier, 2007; Vendetti & Bunge, 2014).

In conclusion, it was possible to detect one specific network sensitive to n-back based cognitive training. Greater connectivity within this network was observed in the training group. The nodes comprised by this network support interference resolution, working memory related processes, intelligence, and motivation. Many of these might help to explain the results reported at the behavioral level. However, note that, to our knowledge, this is the first study that focused on changes in structural connectivity after a cognitive training program. Therefore, more studies are necessary to confirm these outcomes.

CHAPTER 9: Follow-up (Q3 & Q4).

9.1. Introduction.

This chapter presents the follow-up results using the framework described in chapter 5. Near-transfer and far-transfer effects are considered. The former is addressed one year after the end of the training program, while the latter is analyzed at different time intervals (3 months, two years and three years after ending the training program).

These are our hypotheses:

(a) Near-transfer effects will be found. This is based on meta-analyses reported by Melby-Lervåg and Hulme (2013) and Schwaighofer et al. (2015), where a moderate effect size (.41) was found. Nevertheless, given the longer time period considered here (1 year vs. 3-6 months), the effect might be smaller.

(b) Far-transfer effects will be absent. The meta-analyses reported by Melby-Lervåg and Hulme (2013) detected small effects in follow-up studies for nonverbal ability (d = -0.06), attention (d = 0.09), decoding (d = 0.13) and arithmetic (d = 0.18). Similar results were found in the meta-analysis reported by Schwaighofer et al., (2015). However, note that we measured academic performance to test these far-transfer effects, and, therefore, this will be the first study verifying the probable social impact of cognitive training (Buschkuehl & Jaeggi, 2010; Jensen, 1998).

9.2. N-back performance: Near-transfer.

The fifty-six participants of the final sample, with both behavioral and brain images, were invited to a voluntary session one-year after the end of the training program. Twenty-two from the training group and fifteen controls accepted the invitation. The near-transfer task was a modified version of the visual adaptive n-back task completed during the training regime (Román et al., 2015). Compared to the original, the stimuli of the modified task: (a) were always displayed in the same place (center of the screen). This meant that whereas the original visual n-back task comprised a spatial component, the modified task did not; and (b) these stimuliwere faces, which are important in social interactions (Frith, 2007). These stimuli make the task more ecologically valid than an abstract n-back task (visual or auditory).

The final sample considered in this follow-up includes thirty-three participants (20 from the training group and 13 controls). There were no statistically significant differences between volunteers and non-volunteers in any cognitive construct measure at the end of the program (fluid intelligence, crystallized intelligence, working memory and attention control). However, the difference was statistically significant in Neuroticism (voluntaries = 19.9 vs. non-voluntaries = 24.8; $t(50) = -2.33$, $p = .023$, $d = -.66$). The differences were not statistically significant in the remaining personality traits considered (Extraversion, Openness to experience, Agreeableness and Conscientiousness).

Firstly, an independent sample t-test was computed to test for group differences in the follow-up n-back task. Performance in the last block was the dependent variable and group (training vs. control) was the independent variable. Cohen *d* was also computed. Results showed a statistical significant difference favouring the training group, $t(31) = 4.095$, $p < .0001$, $d = 1.59$. Afterwards, a 12x2 ANOVA (Block x Group) was computed. Block was a repeated measured factor, while group was a between group factor. The Greenhouse-Geisser correction was applied when the sphericity assumption was violated. Partial eta squared (η_p^2) was computed as measure of effect size and Bonferroni correction was applied for multiple comparisons. Significance level was set at $p < .05$. Figure 9.1 shows the results.

Main effects for Block, $F(11,341) = 74.883$, $p < .001$, $\eta_p^2 = .707$ and Group, $F(1,31) = 24.484$, $p < .001$, $\eta_p^2 = .441$ were observed. The interaction was also statistically significant, $F(11,341) = 8.161$, $p < .001$, $\eta_p^2 = .208$. Post-hoc analysis indicated that group differences were present from the third block onwards. Remember that the adaptive nature of the task means that the level of difficulty for each block is related with previous performance.

Differences among top performers, standard performers, and controls were also analyzed. A one-way ANOVA was computed: performance in the last block was the dependent variable and group (top, standard and control) was the independent variable. Partial eta squared (η_p^2) was computed as measure of effect size and Bonferroni correction was applied for multiple comparisons.

Figure 9.1. Performance in faces n-back task for training (green) and control (red) groups. ** = Significant differences between groups. B = block.

A main effect of Group was found, $F(2,30) = 14.210$, $p < .001$, $\eta_p^2 = .486$. A post-hoc analysis showed that performance was different for the three groups: top > standard > control. Performance across blocks was also analyzed by computing a 12x2 ANOVA (Block x Group). Block was a repeated measured factor, while Group (top, standard and control) was a between factor. The Greenhouse-Geisser correction was applied when the sphericity assumption was violated. Partial eta squared (η_p^2) was computed as measure of effect size and Bonferroni correction was applied for multiple comparisons. Significance level was set at $p < .05$. Figure 9.2 shows the results.

All effects (Block, Group, Interaction) were statistically significant: [$F(11,330) = 103.077$, $p < .001$, $\eta_p^2 = .775$; $F(2,30) = 15.089$, $p < .001$, $\eta_p^2 = .501$; $F(11,330) = 6.386$, $p < .001$, $\eta_p^2 = .299$ respectively]. Post-hoc analyses showed that (a) controls' performance was lower from block 3 onwards, and (b) top and standard performers showed remarkable performance differences in the latest (more difficult) blocks only.

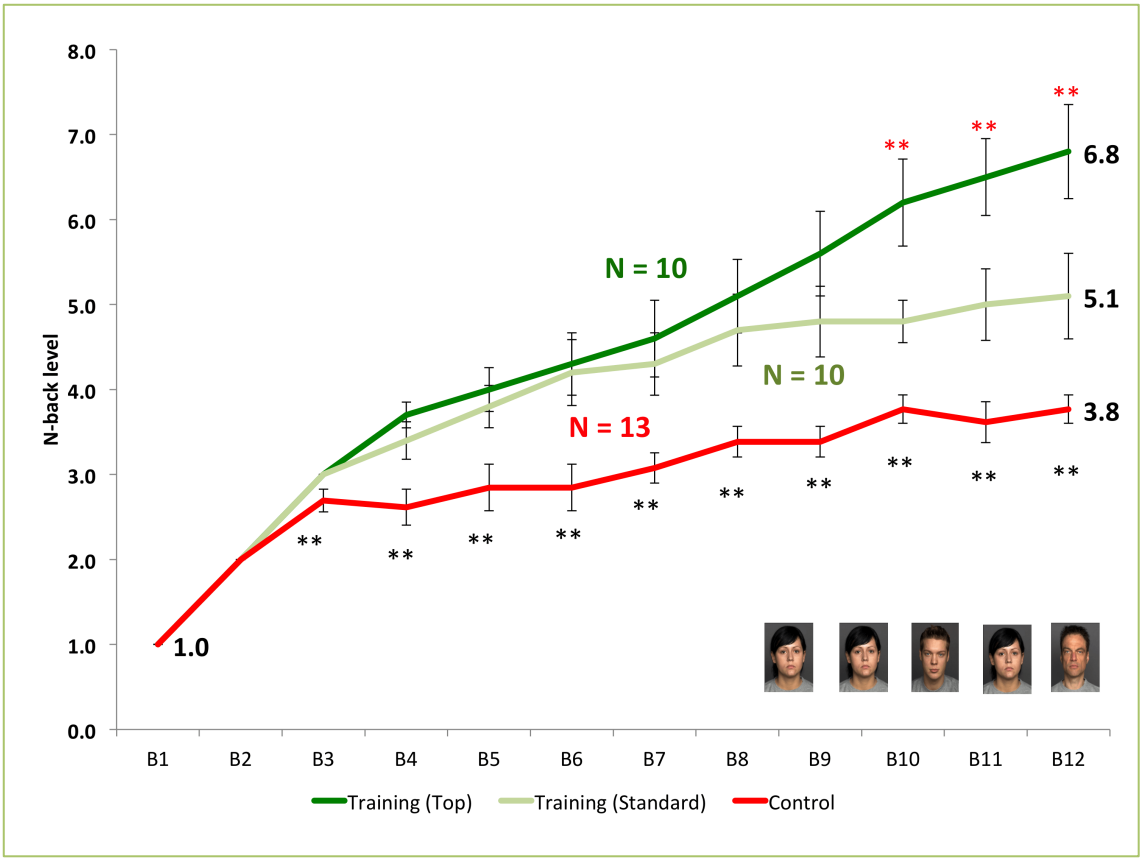


Figure 9.2. Performance in faces n-back task for top training (dark green), standard training (green) and control (red) groups. ** = Significant differences between training (top and standard) and control groups, ** = Significant differences between top and standard training groups. B = block.

Finally, we checked the performance of the training group with their performance in the very first session completed a year beforehand. This result was compared with performance of the control group in the follow-up task. These analyses were computed to detect if the performance for both groups was different the first time that they completed the n-back. Figure 9.3 shows the results.

A one-way ANOVA was computed to check if controls' performance in the last block of the follow-up task differed from performance of the training group one year before. Interestingly, no differences were found, $F(1,31) = 1.228$, $p = .276$, $\eta_p^2 = .083$.

In summary, we conclude that performance of the training group was much better in the follow-up task, one year after completing the training program, than controls' performance. These differences were also revealed for top performers, standard performers, and controls.

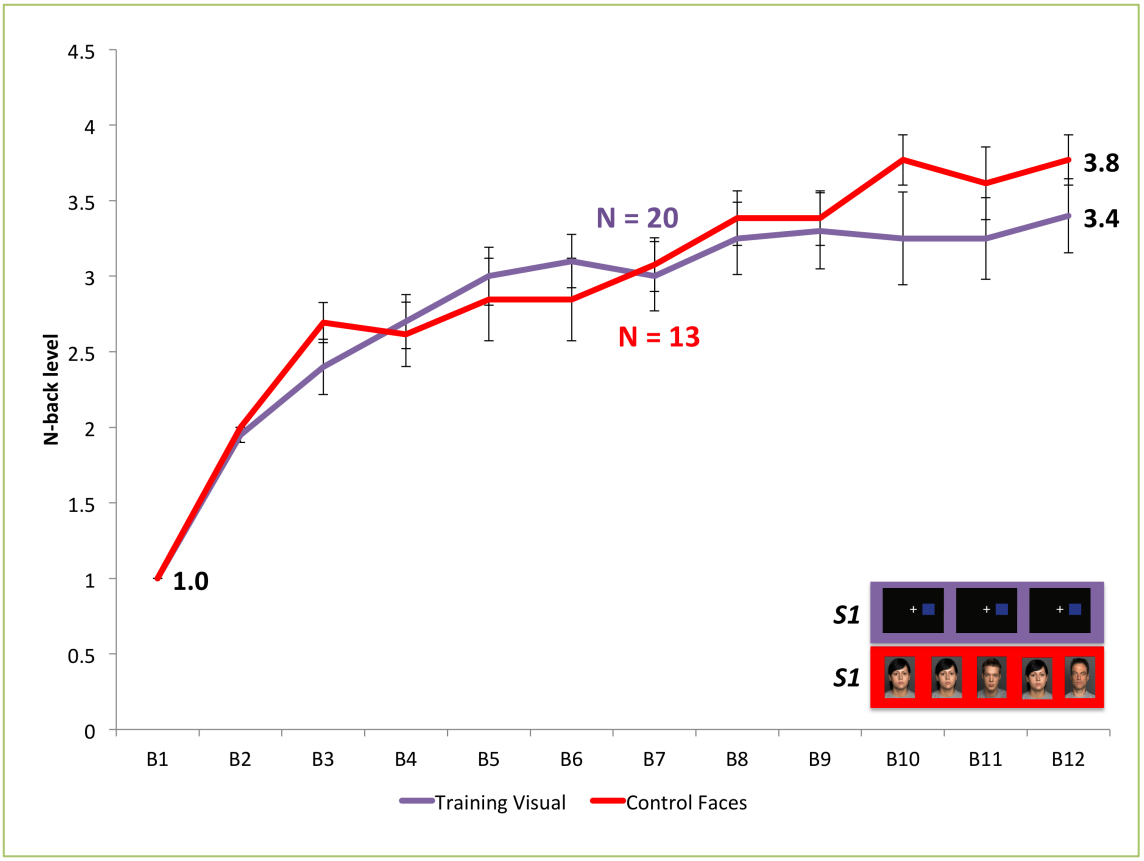


Figure 9.3. Performance in faces n-back task for control (red) group, and for training group in session one of visual n-back task one year before (violet). B = block. S = session.

9.3. Academic performance: Far-transfer.

Academic performance on Statistics, Psychometrics, and Psychological treatment was measured (a) 3 months after the training program (Statistics), (b) 2 years after the training program (Psychometrics), and (c) 3 years after the training program (Psychological treatments). For these academic subjects, participants completed the same standardized knowledge test: Statistics (N = 52), Psychometrics (N = 43), and Psychological treatments (N = 40).

Correlations between psychological factors at the post-test and academic grades are shown in Table 9.1.

Note that academic performance showed a positive manifold, with correlation ranging from .28 to .59. Fluid and crystallized intelligence showed medium effect sizes ($r > .25$) with Statistics and Psychometrics. Psychological treatments' grades were related to crystallized intelligence only. Working memory and attention control were not

correlated with academic performance. These results suggest that Statistics and Psychometrics would reveal group differences (training vs. controls).

Table 9.1. Correlations between psychological factors and academic grades. Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity, ATT = Attention control. Psychological measures were assessed after n-back training program.

	Statistics	Psychometrics	Psychological treatments	Gf	Gc	WMC	ATT
Statistics	1	.59	.28	.26	.29	.03	-.08
Psychometrics		1	.35	.30	.27	.13	.07
Psychological treatments			1	.02	.23	.12	-.02

Table 9.2 shows differences in academic performance for the training and control groups (t student and effect size)

Table 9.2. Performance on academic subjects for the training and control groups. df = degree freedom.

	Training	Control	*t(df), p (one-tail)*	*d*
Statistics	4.72 (1.69)	5.29 (2.13)	-1.06(50), $p = .14$	-0.30
Psychometrics	6.17 (1.62)	6.59 (2.35)	-0.67(41), $p = .25$	-0.21
Psychological treatments	6.65 (1.58)	7.26 (1.31)	-1.31(38), $p = .09$	-0.42

No statistical significant differences were found. Therefore, far-transfer effects were not observed. Indeed, academic performance of the control group was higher.

9.4. Discussion.

9.4.1. *Near-transfer.*

The near-transfer effect reported here was very robust: $d = 1.59$ vs. $d = .41$ (Melby-Lervåg & Hulme, 2013; Schwaighofer et al., 2015). Group differences in the follow-up task were revealed in the very first block of trials (see Figure 9.1). Furthermore, achievement levels observed in top performers reinforced the argument related with the efficiency hypothesis discussed in chapter 6.

The training group showed higher performance than controls one year after completing the training program, which supports the conclusion that training effects remain in the long run. Correlations with fluid intelligence and with working memory capacity also

remain one year later, which, again, suggest that available cognitive resources at baseline play a substantial role.

However, our near-transfer measure is closely similar to the original training task, which can be seen as a limitation. We share this limitation with previous studies, such as the reported by Dahlin, Nyberg, Bäckman, and Neely (2008), who administered a verbal working memory task in their follow-up very similar to their training tasks. Therefore, we cannot conclude that the advantage observed for the training group would be generalized to further measures tapping working memory processes, such as the reading span, dot-matrix, or computation span tasks. As noted by Jaeggi et al. (2010a,b) or Schmiedek et al. (2009), cognitive processes tapped by the n-back task do not show a large overlap with other working memory complex tasks.

9.4.2. *Far-transfer.*

Far-transfer effects are elusive (Melby-Lervåg & Hulme, 2013; Schwaighofer et al., 2015; Jaeggi et al., 2011, 2014; Li et al., 2008; Thompson et al., 2013). Here we chose to measure these effects by academic grades, mainly because it had not been done before (Buschkuehl & Jaeggi, 2010; Estrada et al., 2015; Hunt, 2012a; Jensen, 1998). The prediction was better scores for the training group in those academic subjects showing correlations with fluid reasoning (Gf).

Expected correlations between grades and Gc for Statistics, Psychometrics, and Psychological treatments were found. Gf showed correlations with Statistics and Psychometrics only (see Table 9.1). Results did not support our prediction (see Table 9.2). Indeed, controls showed a systematically better academic performance than the training group. Note that this difference cannot be attributed to probable IQ differences because both groups were matched on this factor before starting the training regime. Therefore, there must be further variables accounting for this discrepancy.

CHAPTER 10: General Discussion.

10.1. Introduction.

The effects of cognitive training programs aimed at increasing working memory related processes has become remarkably controversial in the past years. Both theoretical reviews (Buschkuehl & Jaeggi, 2010; Buschkuehl et al., 2013; Conway & Getz, 2010; Haier, 2014; Moody, 2009; Shipstead et al., 2012) and meta-analyses (Au et al., 2014; Melby-Lervåg & Hulme, 2013; Schwaighofer et al., 2015) are now available and they may help to achieve a relatively solid evaluation of the evidence.

The present doctoral dissertation has been focused on a set of issues closely related with this training research:

Q1: Training performance and transfer effects.

Q2: Structural changes in the brain after training.

Q3: Long-term maintenance of training achievement.

Q4: Transfer to academic performance.

This final chapter summarizes and discusses answers to these questions provided by the completed research. It is important to underscore that our recruitment procedure followed a careful matching process unseen before (Au et al., 2014). Training and control groups were matched according to their intelligence scores maximizing heterogeneity/variability. Moreover, only women were recruited in order to avoid potential confounds related to acknowledged sex differences in brain structure (Escorial et al., 2015; Ruigrok et al., 2014; Tang et al., 2010).

10.2. Q1: Cognitive training and Transfer effects.

The cognitive training program was based on the adaptive dual n-back task designed by Jaeggi and colleagues (Jaeggi et al., 2008). They followed Halford et al.'s (2007) framework, which underscored the shared capacity limitations of fluid reasoning and working memory capacity: "*these limits are quantified in terms of the number of items that can be kept active in working memory, and the number of inter-relationships*

between elements that can be kept active in reasoning" The latter number defines the complexity of typical problems comprised in fluid reasoning standardized tests (see Colom et al., 2005, 2006, 2007). From this perspective, overcoming these capacity limitations by completing (successfully) properly designed cognitive training programs would contribute to improve both psychological factors (Jaeggi et al., 2008).

Note that this task is usually employed in neuroimaging studies (Conway et al., 2005) and the meta-analysis by Au et al. (2014) reported a far-transfer effect after n-back training with healthy young adults. The conclusion is much weaker for a wider set of cognitive programs based on tasks tapping various working memory related processes (Melby-Lervåg & Hulme, 2013; Schwaighofer et al., 2015). The present research endorsed key recommendations for these sorts of short-term cognitive training programs (Buschkuehl, & Jaeggi, 2010; Jaeggi et al., 2010a; Klingberg, 2010; Morrison & Chein, 2011).

These were the main results regarding n-back training:

(a) Large improvements were found across sessions. Specifically, the mean performance achieved in last session of the dual n-back (session sixteen) was 5.2. Table 10.1 shows where this result is located among published studies.

Note that transfer effects were only found in studies with high improvements in the training program. Practice sessions based on single visual and auditory n-back, along with the use of systematic feedback, may be related with this success. In this regard, Jaeggi et al. (2014) showed their participants a curve representing their performance with respect to a generic curve computed after previous data collected in their laboratory. In our study, a direct comparison between participants' performance and one comparable American sample was presented in session eight of the dual n-back training. The disappointing results observed by Chooi and Thompson (2012), Heinzel et al. (2014), Redick et al. (2012), Salminen et al. (2012) or Thompson et al. (2013) might be accounted for, at least in part, by the low achievement levels of their participants on the training regime (Table 10.1). Therefore, we suggest that training improvements might be substantially related with transfer effects.

Table 10.1. Achieved N-back performance in a set of studies.

Study	N training	Number of sessions	Back in last session	Far Transfer
Schweizer et al. (2011)	29	19	7	Satisfactory
Jaeggi et al. (2010)	25	20	7	Satisfactory
Jaeggi et al. (2014)	25	20	6.2	Satisfactory
Jaeggi et al. (2008)	15	19	5	Satisfactory
Salminen et al. (2012)	20	16	4.9	Null
Thompson et al. (2013)	20	20	4.8	Null
Redick et al. (2012)	10	20	4	Null
Chooi & Thompson (2012)	13	20	3.8	Null
Heinzel et al. (2014)	15	12	4	Null
Stephenson & Halpern (2013)*	28	18	< Jaeggi et al. (2008)	Satisfactory
Mean	20	18.4	5.2	50% Satisfactory

* Stephenson & Halpern (2013) did not provide precise information about this issue.

(b) Performance differences across training sessions correlated with baseline individual differences in fluid intelligence (Gf), crystallized intelligence (Gc) and working memory capacity (WMC), but not with attention control (ATT) and personality traits. The n-back task was designed to measure working memory related processes (Conway et al., 2005). The correlation between Gf and WMC supports that both psychological factors share common cognitive processes. The sustained correlation values between Gf, WMC, and n-back performance across sessions supports the adaptive nature of the training program (Buschkuehl, & Jaeggi, 2010; Jaeggi et al., 2010a; Morrison & Chein, 2011, see von Bastian & Eschen, 2015). Gf and Gc were highly correlated at the latent level (r = .95) in the reference sample (N = 128). This very high correlation is interesting because, as described below, transfer effects for Gf and Gc were completely different.

In short, the cognitive training program worked satisfactorily, and, therefore, analysis of the probable effects at both the psychological and biological levels were worthwhile.

In this regard, near-transfer effects at the construct level (WMC) were not statistically significant. Far-transfer effects at the construct level (Gf, Gc and ATT) were also absent, consistent with previous reports using similar research approaches (Chooi & Thompson, 2012; Redick et al., 2012).

Far-transfer effects are difficult to explain in the absence of near-transfer effects (Shipstead et al., 2012). However, n-back related cognitive processes and working memory related processes show partial overlap (Jaeggi et al., 2010a,b; Kane et al., 2007; Schmiedek et al., 2009). Jaeggi et al., (2008) suggested that attention control is shared by Gf measures and the n-back task. However, we also failed to find any effect over attention control at the construct level.

Colom et al. (2013a) highlighted the acknowledged fact that constructs are not homogeneous entities. Therefore, looking at specific tests and tasks is relevant. These were the main results:

(i) The training group obtained greater changes than controls in all Gf measures. Results for the RAPM showed a remarkable group difference (the standardized change for controls was -.10, while for the training group it was .18). The RAPM requires substantial visual-spatial processing skills in addition to rule management and fluid reasoning (Abad, Colom, Rebollo & Escorial, 2004; Carpenter, Just & Shell, 1990).

(ii) The standardized changes observed for the three measures of crystallized intelligence (Gc) were null for both groups. Note that this result departs from the Hawthorne effect taken for granted in this kind of research (Shipstead et al., 2012).

(iii) We found statistically significant group changes in two out of three working memory measures. Dot matrix and reading span improvements were substantially greater for the training group (training group showed an increment of .23 and .90 for reading span and dot-matrix respectively, while for the control group these values were .07 and .58, respectively). The opposite result was observed for the computation span task (the control group showed an increment of .55, while the increment was .22 for the training group), which helps to explain the null finding at the construct level for working memory. We computed a WMC factor using the reading span and dot matrix tasks and the results revealed a standardized change of .46 for the training group and of .26 for controls. The ANCOVA analysis showed that these differences were statistically significant, $F(1,53) = 4.921$, $p = .015$, $\eta_p^2 = .085$.

(iv) The Simon task, which measures spatial attention, revealed a statistical significant difference (standardized change for the training group was .15, while the outcome was a decrement in performance in controls, -.17).

In conclusion, post-hoc analysis showed that visual-spatial processing skills, as well as the resolution of interference in spatial tasks, across cognitive domains, might be sensitive to the completed training (Jaeggi et al., 2014; Melby-Lervåg & Hulme, 2012; Hindin & Zelinski, 2012; Uttal et al., 2013). Stephenson and Halpern (2013) tested directly this issue by dividing their sample in four training groups: (a) dual n-back, (b) visual n-back, (c) auditory n-back and (d) short-term (STM) visual task (a passive control group was also considered). Their main question was: is the type of training differentially related with Gf? People who trained with spatial tasks (dual n-back, visual n-back and STM visual task) obtained higher transfer effects than controls. However, these differences were only marginally significant for the auditory vs. control comparison. Therefore, the spatial component might be more relevant than the WMC component to obtain satisfactory transfer effects. Furthermore, they argued that STM training enhanced the shared short-term storage component underlying Gf, which was seen as consistent with reports analyzing the limitations shared by working memory and cognitive ability (Colom et al., 2006, Martínez et al., 2011, Krumm et al., 2009, Hornung et al., 2011). Generally speaking, people have less practice on spatial than on verbal tasks, and, therefore, there may be more room for improvement in the former (Miyake et al., 2001).

10.3. Q2: Structural Brain Changes after Cognitive Training.

N-back training shares a common capacity limit with fluid intelligence. Therefore, improvements in working memory related processes may impact on fluid intelligence (Jaeggi et al., 2008). The finding that Gf and WMC share a common brain network reinforces this perspective (Barbey et al., 2014a; Colom et al., 2007; Gray et al., 2003). In this regard, Colom et al. (2007) found that both factors share a common anatomic framework located mainly in frontal grey matter (right superior frontal gyrus and left middle frontal gyrus) and in the right inferior parietal lobule. Similar results were found at the functional level (Gray et al., 2003) and with lesion patients (Barbey et al., 2014a).

Importantly, Barbey et al. (2014a) detected high structural overlaps between Gf and spatial WMC, verbal/numerical WMC and manipulation. However, the overlap between the monitoring factor (defined with three levels of difficulty on the n-back task) and Gf was very small, which was consistent with the low correlation at the latent variable level (-.16). This may help to explain the lack of transfer at the construct level observed here for Gf and WMC.

The average correlation in our study between Gf and n-back performance was $r > .40$ across training sessions, which underscores the relevance of the adaptive nature of our n-back task (Jaeggi et al., 2010b). This result increases the confidence in finding some brain responsiveness to training (along with transfer effects to related psychological factors). In order to examine this issue, we analyzed the time x group interaction, as recommended by Thomas and Baker (2013a).

Three structural brain indices were analyzed: (a) cortical thickness (CT), (b) cortical surface area (CSA), and (c) structural connectivity (SC).

Firstly, we computed the correlation between n-back performance and these indices at the vertex and ROI levels. Results were weak. Specifically, variations in cortical surface area in frontal, parietal and temporal regions (dorsal and ventromedial frontal cortex, motor cortex, precuneus, superior-middle temporal and part of inferior parietal cortex) were related with n-back performance (range .20 - .50), while correlations for cortical thickness were located mainly in posterior regions (occipital lobe, middle and inferior temporal cortex and parietal lobe) with a range of 20 - .35. Connectivity measures at different levels of analysis (nodal and global) were unrelated with training performance. Note that findings for surface area were consistent with previous research (Barbey et al., 2014a; Tsuchida & Fellows, 2009).

Secondly, we computed standardized changes (Jaeggi et al., 2011) for the biological indices and we analyzed the time x group interaction. Statistically significant group differences were found in several regions of interest (ROIs) for cortical thickness (CT) and cortical surface area (CSA). A summary of these results is shown in Figure 10.1. Regions where the training group showed thickening or preservation, along with expanding or CSA preservation with respect to controls were located in right ventral

frontal cortex (CT), right pars opercularis (CSA), right middle temporal cortex (CT and CSA), and one small region in the right inferior parietal region.

The research strategies for analyzing changes in structural connectivity [(a) node level, (b) global level, and (c) network analysis based on network-based statistics, NBS] showed statistical significant results for the interaction at the sub-network level only: the training group showed an increment in generalized fractional anisotropy (GFA). This network comprised subcortical (bilateral forebrain, left pallidum, right accumbens), frontal (right pars triangularis and right pars opercularis), parietal (left supramarginal, left inferior parietal and right postcentral), and temporal regions (left parahippocampal, and left middle temporal), along with the right insula. These nodes support (a) inhibition, (b) working memory-intelligence, and (c) engagement. The most central node (highest connected with the remaining nodes) within this sub-network network was located in the middle temporal lobe (see Figure 10.2).

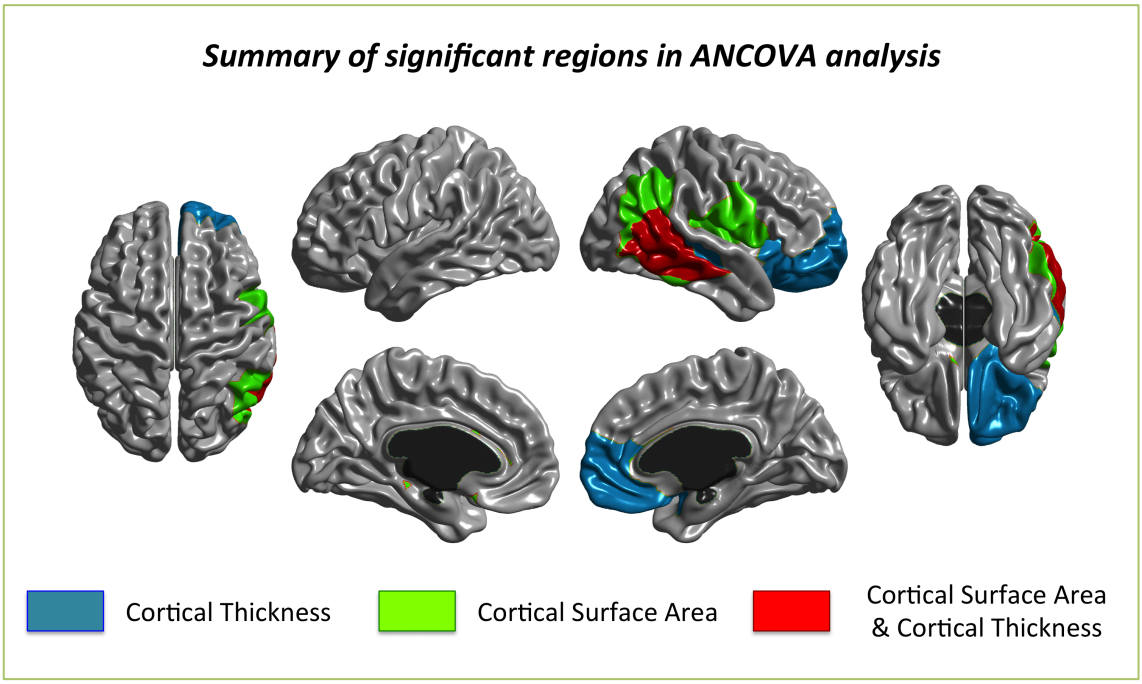


Figure 10.1. Summary of statistically significant results after the interaction analysis time x group for cortical thickness and cortical surface area.

In summary, we conclude that findings from surface-based morphometry (CT and CSA) and structural connectivity analyses suggest that the middle temporal region is the brain region most sensitive to the completed training. Previous research has shown that this brain region supports working memory-related processes (Hampson, et al., 2006; Zou et

al., 2013) and interference resolution (Kirwan, & Stark, 2007; Yassa, Mattfeld, Stark, & Stark, 2011; Yassa & Stark, 2011) highly relevant in coping with the n-back task (Jaeggi et al., 2010b). The recent meta-analysis by Basten et al. (2015) underscored the relevance of brain structural differences in the temporal lobes in supporting high-level cognitive processes. The role of the parietal lobes seems more relevant when brain functional differences are considered (see also Lange et al., 2010).

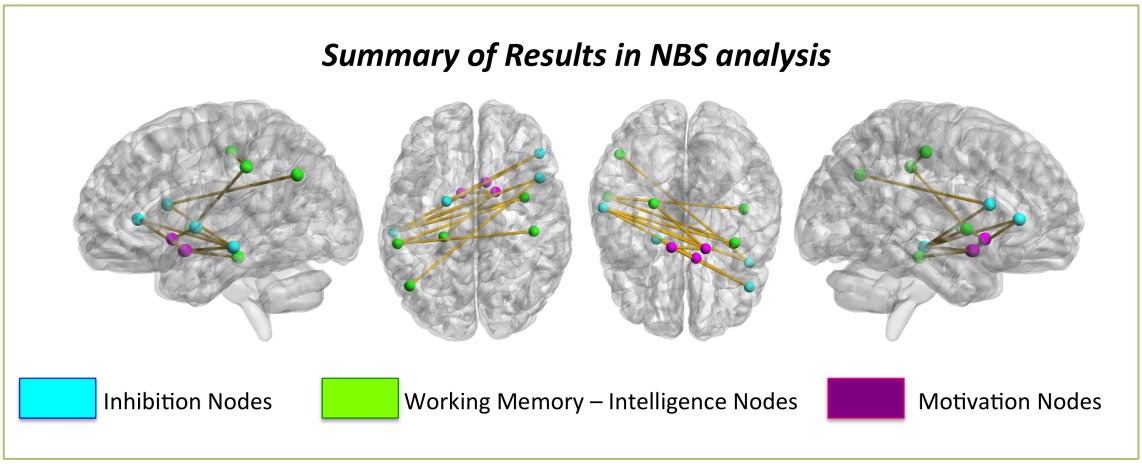


Figure 10.2. Summary of statistically significant results in the NBS analysis.

Different biological mechanism might help to explain the changes found for the analysed grey and white matter indices. For example, Zatorre et al., (2012) considered that axon sprouting, dendritic branching and synaptogenesis, neurogenesis, changes in glial number and morphology, and angiogenesis could be the cellular mechanisms that explain the changes detected by MRI. On the other hand, changes in white matter could be explained by alterations in fiber organization, which could include axon branching, sprouting, packing density, axon diameter, fiber crossing and the number of axons; myelination of unmyelinated axons; changes in myelin thickness and morphology; changes in astrocyte morphology or number; and angiogenesis. Unfortunately, the current spatial resolution of the MRI images precludes finding an answer to these possibilities. Therefore, studies analysing neurotransmitters may help to understand the mechanics underlying training and their effects over psychological traits such as intelligence. In this regard, McNab et al. (2009) found that the cortical dopamine D1 receptor binds potential changes in prefrontal and parietal brain regions after working memory training.

10.4. Q3: Training performance one year later.

Several participants from the training and control groups accepted the invitation to participate in a follow-up session that took place one year after ending the training program. They completed an adaptive session using a variation of the single visual n-back task (Román et al., 2015). The group performance difference favouring the training group was large (d = 1.59) and it was observed from the very first trials.

This result was greater than those reported in published meta-analyses (Melby-Lervåg & Hulme, 2013; Schwaighofer et al., 2015), but this might be explained by the similarity between the original and follow-up tasks (see also Dahlin et al., 2008a).

10.5. Q4: Academic performance differences.

We also measured academic grades as a far-transfer index (Buschkuehl, & Jaeggi, 2010; Estrada et al., 2015; Hunt, 2012a; Jensen, 1998). The academic grades were collected at different time points and for different academic topics: (a) statistics (3 months after the training program), (b) psychometrics (2 years after the training program), and (c) psychological treatments (3 years after the training program). We chose these course topics because participants completed the same standardized academic knowledge test, i.e. there were not differences in the topics of the courses between students from different classrooms. The correlation between academic grades was within the range .28 -- .59 (see Figure 10.3).

We failed to find any statistical significant group difference. Indeed, controls showed better academic scores than the training group. Note that this average difference cannot be attributed to baseline IQ differences because both groups were carefully matched in this regard. Indeed, the correlation between intelligence and academic grades for our groups were largely comparable to those usually observed for university students (around 0.30; Jensen, 1998).

Looking at the individual level, we observed that three students of the control group obtained very high academic scores (mean in all academic courses higher or equal to 8.5). The reanalysis excluding these three students reduced the group difference (for obvious reasons). Specifically, for statistics, the mean was of 4.7 and 4.8 for the training and control groups respectively. For psychometrics, the mean was equal in both groups

(6.2), and for psychological treatment, the mean was 6.7 for the training group and 7.0 for controls.

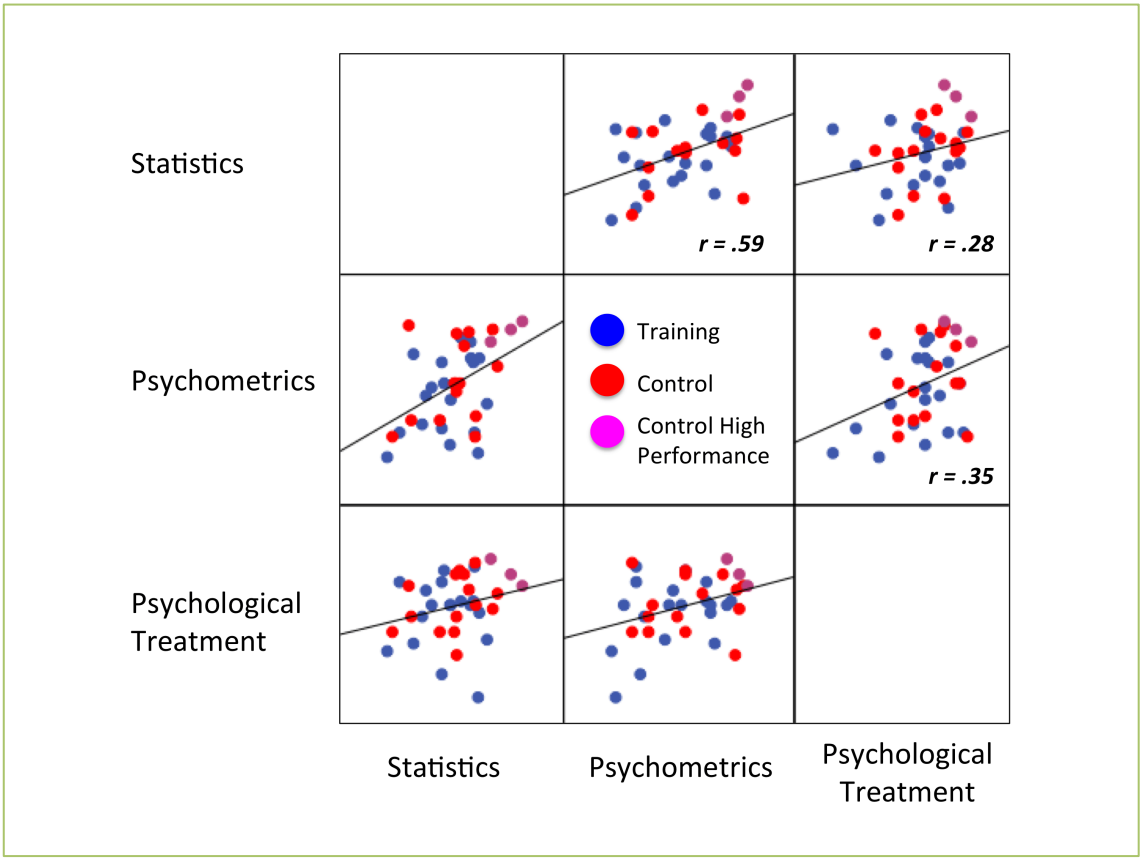


Figure 10.3. Scatterplot for different academic grades.

10.6. Top and standard performers on the training program.

Dividing the training group in top and low performers has been done before within the working memory training research for testing transfer effects (Jaeggi et al., 2011; Kundu et al., 2013; Redick et al. 2013; Rudebeck et al., 2012; Thompson et al. 2013). The researcher may define a threshold on an outcome variable, or performance groups are defined on one variable dichotomizing difference scores between a first and last n-back training session with a median split (Tidwell, Dougherty, Chrabaszcz, Thomas, & Mendoza, 2014).

However, Tidwell et al., (2014) published a critical theoretical review about this practice. They concluded that it is inappropriate to test hypotheses regarding the effectiveness of cognitive training, since the division of the training group according to an arbitrary criterion is an artifact. Indeed, they implemented a series of simulations and

demonstrated that the apparent transfer effects observed by dividing the groups are illusory and independent of the effectiveness of cognitive training. This is so because there are positive correlations between changes in the training and changes in the psychological measure. Another limitation is the lower power as a result of categorizing a continuous variable. Note, however, that changes in the training and changes in the psychological measure were uncorrelated in our study ($r_{Change\_Gf\text{-}Change\_training}$ = -.093, p = .651; $r_{Change\_Gc\text{-}Change\_training}$ = .014, p = .946; $r_{Change\_WMC\text{-}Change\_training}$ = -.090, p = .662 and $r_{Change\_ATT\text{-}Change\_training}$ = .250, p = .218).

We divided the group using training level in last session of the dual n-back task, obtaining two groups: (a) top performers (N = 10; n-back level achieved higher than 5.2) and (b) standard performers (N = 16; n-back level achieved lower than 5.2).

Following this strategy, we obtained the results described below.

10.6.1. *Psychological measures*

(a) Top and standard performers differed from the outset (before training) in fluid intelligence (Gf), crystallized intelligence (Gc) and working memory capacity (WMC), but not in attention (ATT) and personality traits. Importantly, top performers found the training task less complex, which supports the relevance of adapting the difficulty level to participants' continuous performance (Jaeggi et al., 2011; Ryan, & Deci, 2000).

(b) A statistically significant effect on Gf was found for top performers, consistent with Jaeggi et al. (2011). Specifically, top performers showed an average standardized increase of .91, whereas standard performers showed an average improvement of .57 (controls showed an average increase of .49).

(c) The training group obtained higher increments than controls in all Gf measures, but this was especially noteworthy for top performers. The largest (and statistically significant) difference among top performers, standard performers, and controls was found in DAT-AR (the standardized change in DAT-AR was .66, .29, .35, respectively). Interestingly, this measure showed the highest factor loading (.74) on the Gf latent factor.

(d) Top performers showed the highest increment in dot-matrix (the standardized change was 1.0, .80 and .60 for top, standard and controls respectively) and in the spatial attention control task (standardized change was .30 for top, .06 for standard and a decrement in performance was found for controls, -.17)

10.6.2. *Biological measures*

The division between top and standard performers on the training program revealed interesting findings at the biological level:

(a) Regarding cortical thickness (CT), top performers and controls obtained a similar pattern of results (preservation or thinning processes), while standard performers showed preservation or thickening. Statistically significant group differences (standard performers > controls) were found in right ventral frontal cortex and right middle temporal.

(b) Results for cortical surface area (CSA) were weaker than for CT. The most relevant changes were found in fronto-temporal regions. Specifically, top performers showed expansions in left posterolateral temporal cortex, whereas standard performers showed expansions in this same region of the right hemisphere. Standard performers showed expansions in left medial temporal cortex and right orbitofrontal cortex, while contractions in left middle temporal cortex were found in top performers. Controls showed expansions in right dorsomedial frontal cortex and CSA contractions in the right temporal area.

(c) For structural connectivity, top and standard performers showed increments in their generalized fractional anisotropy (GFA) in comparison with controls. A very similar network was found and no differences between standard and top performers were found. The main differences were observed for the right supramarginal and right precentral regions, since these nodes were detected in the comparison top performers > controls, but not in the comparison standard performers > controls. The relevance of the connections between these regions for higher-order mental processes has been underscored by Vendetti & Bunge (2014). This result may support the greater transfer effects observed in the top performers.

10.6.3. *Transfer effects and brain efficiency*

We argued above that results for top and standard performers might support the brain efficiency hypothesis (Haier, 1993; Haier et al., 1988; Neubauer & Fink, 2009). The scatterplots showing the relationship between Gf and biological changes in ROIs where the training group showed preservation or increment with respect to the control group were depicted in Figure 10.1.

Figure 10.4 (top panel) shows the relationship between changes in statistically significant ROIs for CT and changes in Gf. This correlation was moderate (r = -.25) in the ventral frontal cortex, and large (r = -.34) in the middle temporal region. This is consistent with the efficiency hypothesis (Tang et al., 2010) and suggests that top performers did have the required processing resources, and, therefore, their CT followed standard developmental changes, whereas standard performers did not. In the latter case, the sustained effortful processing across sessions might be behind the observed thickening.

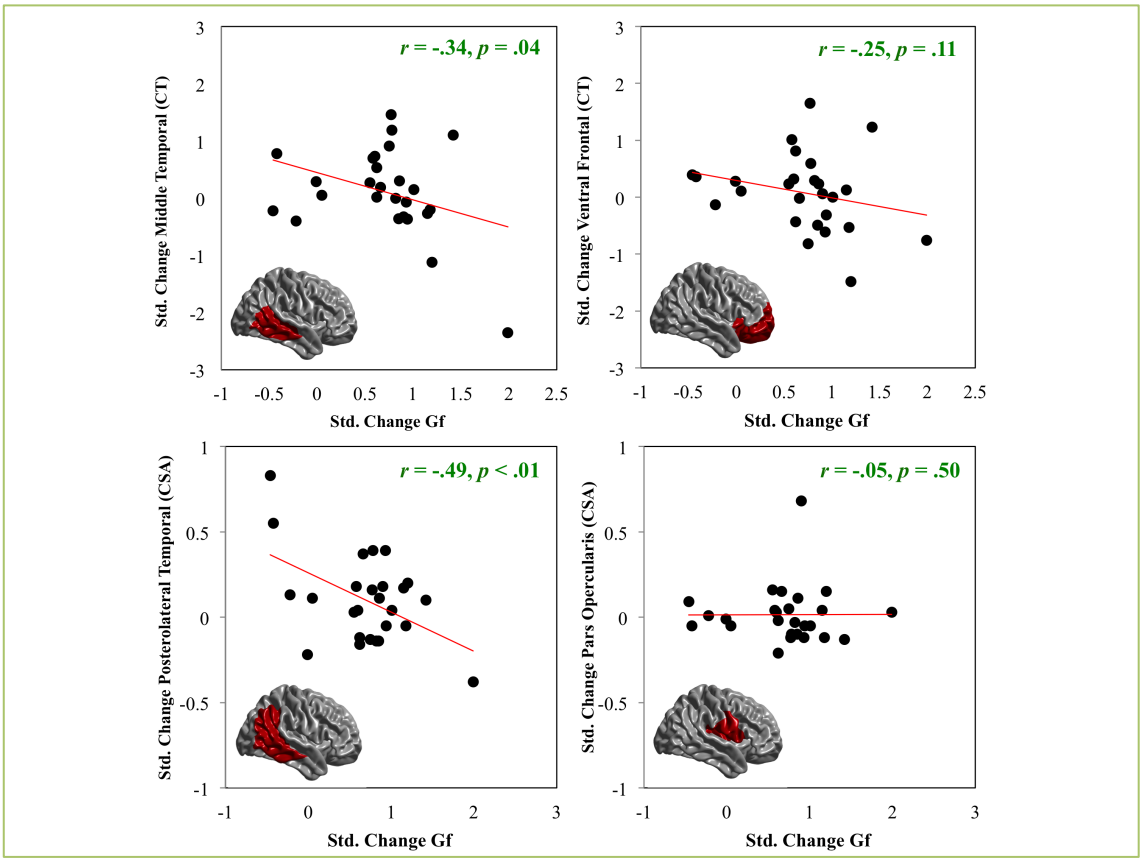


Figure 10.4. Correlations between standardized change in cortical thickness ROIs (top panel) and cortical surface area ROIs (bottom panel) and standardized change in fluid intelligence factor. P value is one-tail.

With respect to CSA, Figure 10.4 (bottom panel) shows the correlations between the standardized changes in two ROIs and Gf. Again, the correlation observed in the middle temporal area (r = -.49) supports the efficiency hypothesis. However, this was not the case for the pars opercularis.

Figure 10.5 shows the scatterplots relating standardized changes in fluid intelligence to changes in connectivity of nodes classified according to their function. The computed correlations were smaller than those observed for cortical thickness and cortical surface area. Nevertheless, values for inhibition nodes (r = -.26) and working memory - intelligence nodes (r = -.21) were moderate.

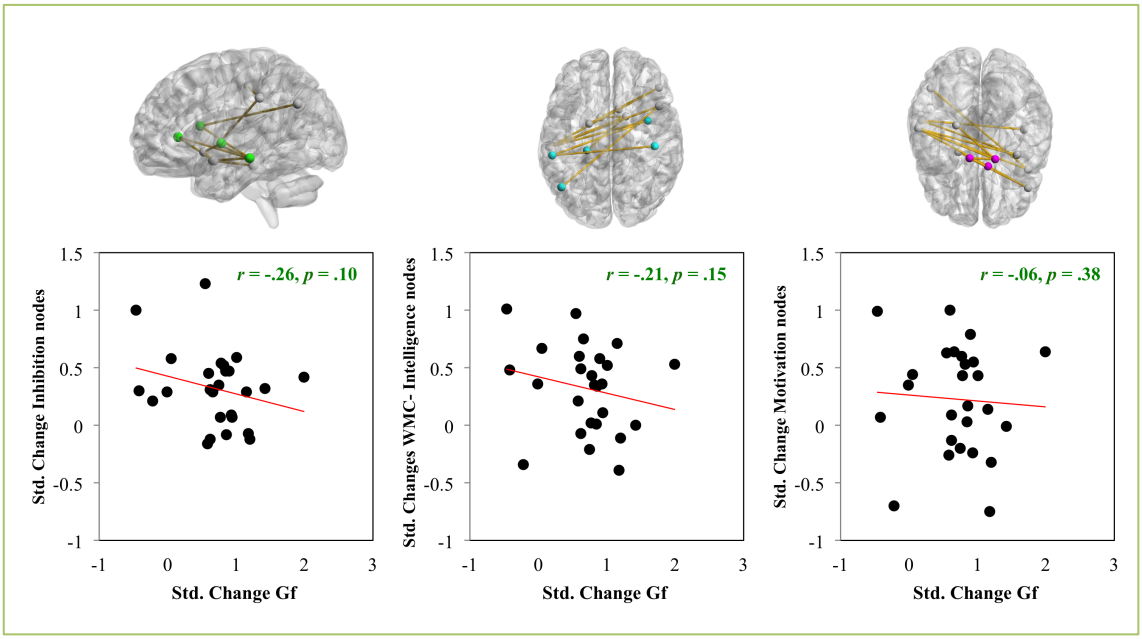


Figure 10.5. Correlations between standardized change in connectivity of different nodes and standardized change in fluid intelligence factor. P value is one-tail.

The brain efficiency hypothesis refers to lower brain activation in high intelligence people during the successful performance of a cognitive task (Neubauer & Fink, 2009). To our knowledge, the efficiency hypothesis has not been addressed before using brain structural data. Animal studies did provide collateral support for our interpretation. Thus, for instance, Curlik and Shors (2011) observed that animals requiring more trials to learn a conditioned response retain more new neurons than animals quickly acquiring the conditioned response. The effort involved in learning is crucial to observe structural brain changes (Shors, 2014).

10.7. Limitations.

There are some possible limitations to our research program.

(a) The sample size employed here is similar to previous behavior studies, but it is low for brain metrics, since the reduction of statistical power is considerable according to the number of statistical tests.

(b) Changes detected in CSA could be mediated by changes in CT. The CSA index is computed using the inner surface (GM-WM boundary). Therefore, changes detected in CT could lead to changes in surface area. Then, a replication for changes detected in CSA with another SBM protocol could be necessary.

(c) Only women were employed, and, therefore, results cannot be directly generalized to men. However, Schmidt et al. (2009) found the same pattern of brain activation during n-back task regardless of the sex variable. Therefore, the structural evidence observed here for women may be equivalent for men.

10.8. Highlights.

These are the main conclusions of the present study.

1. We observed the required improvement levels in the completed training program.

2. The training program had a statistically significant effect over improvements in fluid intelligence at the construct level in top performers only.

3. Analyses at the measures level suggest that visuo-spatial skills tapped across cognitive domains are sensitive to the completed training program.

4. Changes in cortical thickness and cortical surface area were primarily focused in temporal and frontal regions thought to support cognitive processes required for successful performance in the completed training regime.

5. Structural connectivity analyses revealed one specific network sensitive to cognitive training. This network supports interference resolution, working memory related processes, intelligence, and engagement/motivation.

6. The brain efficiency hypothesis was partially supported.

7. Near-transfer effect (performance in faces n-back task) was observed in the follow-up study, but far-transfer (academic performance) was absent.

Conclusiones

"El cerebro no es del todo como un ordenador sino más bien como un ordenador que está cambiando continuamente."

Roger Penrose (1989)
La nueva mente del emperador.

Esta tesis doctoral se ha centrado en el análisis exhaustivo de los efectos psicológicos y biológicos de un exigente entrenamiento cognitivo usando un clásico diseño pretest-posttest en el que se ha comparado un grupo experimental con un grupo control.

El programa de entrenamiento estuvo basado en la tarea n-back (Jaeggi y cols., 2008) porque la evidencia previa sugería que la mejora experimentada durante el entrenamiento se asociaba a un efectivo positivo sobre la inteligencia fluida (Gf).

Los factores psicológicos medidos en el pretest y en el posttest ha sido la inteligencia fluida (Gf), la inteligencia cristalizada (Gc), la memoria operativa (WMC) y control atencional (ATT). A nivel biológico se han valorado los cambios estructurales en el cerebro que pudieran ser atribuidos al entrenamiento cognitivo. En la Figura 1 se presenta de forma esquemática el diseño planteado en la investigación.

Finalmente, se hizo un seguimiento un año después de haber finalizado el programa de entrenamiento. En este estudio se midió el rendimiento a participantes de ambos grupos en una tarea "n-back" similar a la empleada durante el entrenamiento (Román y cols., 2015). Además, se consideraron variables como el rendimiento académico en diversos momentos temporales.

Siguiendo las recomendaciones de Thomas y Baker (2013) se analizó la interacción Tiempo x Grupo.

Estos han sido los principales resultados:

1. Los individuos que completaron el entrenamiento mostraron mejoras similares (o superiores) a las observadas en estudios previos. La correlación de las diferencias de

rendimiento a través de las sesiones de entrenamiento con las medidas psicológicas de Gf, Gc y WM se mantuvo durante todo el programa, hecho que demuestra la ausencia de automatización. El programa resulta cognitivamente exigente de principio a fin.

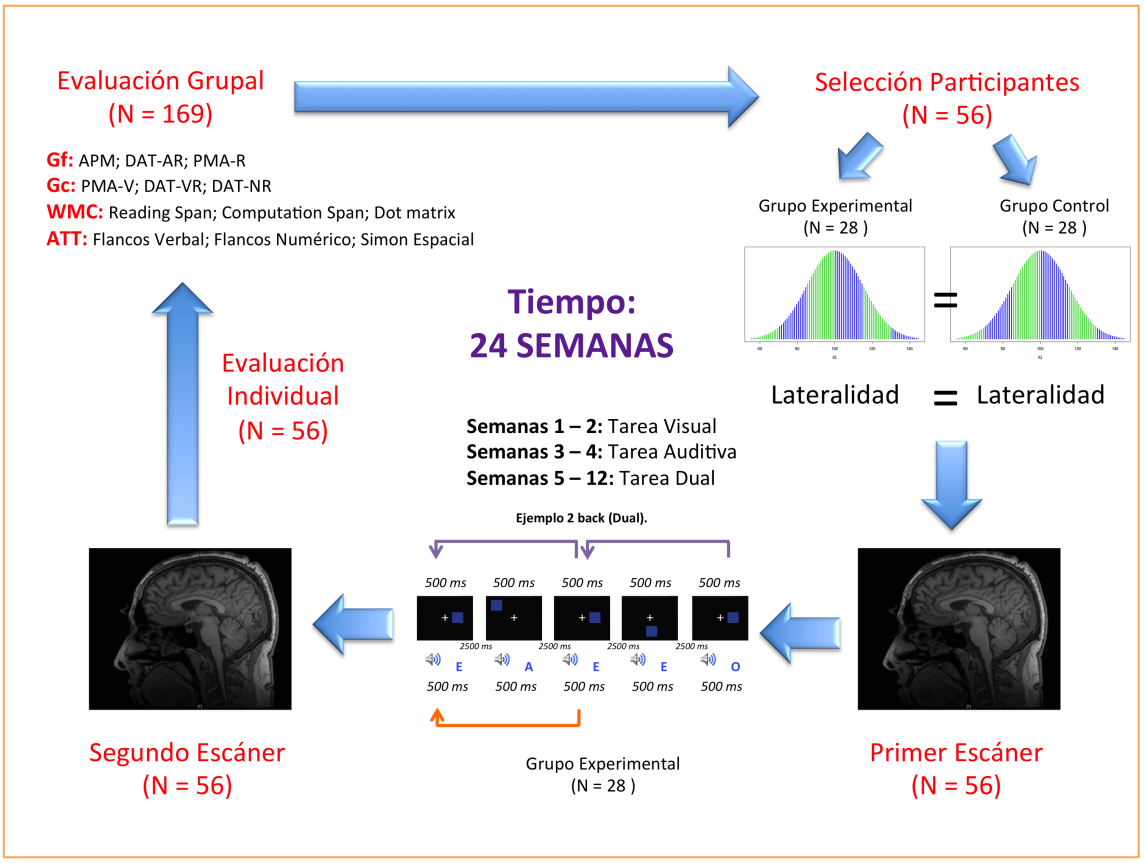


Figura 1. Diseño de investigación. Gf = inteligencia fluida; Gc = inteligencia cristalizada: WMC = memoria operativa; ATT = control atencional. CI = cociente intelectual.

2. No se observó transferencia positiva (near & far transfer) en los constructos evaluados en el pretest y en el posttest. Sin embargo, el análisis del cambio a nivel de test y tarea (medida) reveló efectos significativos en las de alto procesamiento viso-espacial, independientemente del dominio cognitivo valorado.

3. A nivel biológico, se detectaron cambios estructurales en zonas temporales y frontales, tanto para el grosor cortical como para el área de superficie. Esas regiones se han vinculado a procesos cognitivos requeridos durante el régimen de entrenamiento, tales como la memoria de trabajo, la inteligencia, la resolución de conflictos.

4. Los análisis de conectividad estructural mostraron mejoras del grupo de entrenamiento (incremento de anisotropía fraccional) en una red asociada a la

inteligencia, la memoria operativa, la resolución de interferencias y la implicación en la tarea (*engagement*).

5. Los cambios observados a nivel biológico se relacionaron con los cambios en inteligencia fluida (Gf), en el sentido de que las personas con mayor cambio en Gf fueron las que menos modificaron sus estructuras cerebrales. Este resultado apoya la hipótesis de eficiencia neuronal cerebral (Neubauer y Fink, 2009)

6. La mejora alcanzada durante el entrenamiento se mantuvo un año después de la finalización del programa.

7. No se encontraron diferencias de rendimiento académico entre el grupo experimental y control.

En suma, tras analizar los resultados de nuestro programa de entrenamiento podemos sugerir que: (a) las capacidades viso-espaciales son modificable después del entrenamiento basado en la tarea de n-back, siendo menos claros los efectos producidos en Gf, (b) se producen cambios estructurales en el cerebro asociados a zonas relevantes para la realización de la tarea y (c) las mejoras de rendimiento logradas durante el entrenamiento son estables a lo largo del tiempo (al menos un año después), aunque (d) por el momento, no se han detectado que estos cambios estén asociadas con un mayor rendimiento académico.

References

Abad, F. J., Colom, R., Rebollo, I., & Escorial, S. (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and individual differences, 36,* 1459-1470.

Abad, F. J., Sorrel, M., A., Román, F. J., & Colom, R. (In press). The Relationships between WAIS-IV Factor Index Scores and Educational Level: A Bifactor Model Approach. *Psychological Assessment*.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30–60.

Ad-Dab'bagh, Y., Lyttelton, O., Muehlboeck, J. S., Lepage, C., Einarson, D., Mok, K., ... & Evans, A. C. (2006, June). The CIVET image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. *In Proceedings of the 12th annual meeting of the organization for human brain mapping* (p. 2266). M. Corbetta.

Amelang, M., & Steinmayr, R. (2006). Is there a validity increment for tests of emotional intelligence in explaining the variance of performance criteria?. *Intelligence, 34,* 459-468.

Anastasi, A. (1934). Practice and variability. *Psychological Monographs*, 45, 5.

Andrés-Pueyo, A. (1996). *Manual de psicología diferencial.* Madrid: McGraw-Hill Interamericana de España.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in cognitive sciences, 8,* 170-177.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry – the methods. *Neuroimage, 11,* 805-821.

Atkinson, R. C. (1974). Teaching children to read using a computer. *American Psychologist*, 29, 169-178.

Atkinson, R. C. (2005). College admissions and the SAT: A personal perspective. *Observer, 18*(5).

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic bulletin & review,* 1-12.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford University Press.

Bai, F., Shu, N., Yuan, Y., Shi, Y., Yu, H., Wu, D., ... & Zhang, Z. (2012). Topologically convergent and divergent structural connectivity patterns between patients with remitted geriatric depression and amnestic mild cognitive impairment. *The Journal of Neuroscience, 32,* 4307-4318.

Baker, C. I., Hutchison, T. L., & Kanwisher, N. (2007). Does the fusiform face area contain subregions highly selective for nonfaces?. *Nature neuroscience, 10,* 3-4.

Barbey, A. K., Colom, R., Paul, E., Forbes, C., Krueger, F., Goldman, D., & Grafman, J. (2014b). Preservation of general intelligence following traumatic brain injury: contributions of the Met66 brain-derived neurotrophic factor. *PloS one, 9,* e88733.

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014a). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure and Function, 219,* 485-494.

Basser, P. J., Mattiello, J., & LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B, 103,* 247-254.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine, 44,* 625-632.

Bassett, D. S., Yang, M., Wymbs, N. F., & Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nature neuroscience, 18,* 744-751.

Basten, U., Hilger, K., & Fiebach, C. J. (2015). Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence, 51*, 10-27.

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., ... & Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic resonance in medicine, 50,* 1077-1088.

Bell, D. S., Harless, C. E., Higa, J. K., Bjork, E. L., Bjork, R. A., Bazargan, M., & Mangione, C. M. (2008). Knowledge retention after an online tutorial: A randomized educational experiment among resident physicians. *Journal of General Internal Medicine, 23,* 1164–1171.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (2005). *Differential Aptitude Test (DAT-5).* Madrid: TEA, S.A.

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin, 107,* 238-46.

Bezzola, L., Mérillat, S., Gaser, C., & Jäncke, L. (2011). Training-induced neural plasticity in golf novices. *The Journal of Neuroscience, 31,* 12444-12448.

Blanton, R. E., Levitt, J. G., Thompson, P. M., Narr, K. L., Capetillo-Cunliffe, L., Nobel, A., ... & Toga, A. W. (2001). Mapping cortical asymmetry and complexity patterns in normal children. *Psychiatry Research: Neuroimaging, 107,* 29-43.

Brans, R. G., Kahn, R. S., Schnack, H. G., van Baal, G. C. M., Posthuma, D., van Haren, N. E., ... & Pol, H. E. H. (2010). Brain plasticity and intellectual ability are influenced by shared genes. *The Journal of Neuroscience, 30,* 5519-5524.

Braver, T. S., & Cohen, J., (2000). On the control of control: the role of dopamine in regulating prefrontal function and working memory. In Monsell, S. and Driver, J. (Eds.), *Attention and Performance XVIII: Control of Cognitive Processing*. MIT Press: Cambridge, MA.

Brinch, C. N., & Galloway, T. A. (2011). *Schooling in adolescence raises IQ*. Oslo, Norway: Research Department of Statistics Norway.

Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry, 15,* 234-238.

Brown, A. L., & Campione, J. C. (1982). Modifying intelligence or modifying cognitive skills: More than a semantic quibble? In D. K. Detterman, & R. J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 215−230). New Jersey, Ablex: Norwood.

Burgaleta, M., & Colom, R. (2008). Short-term storage and mental speed account for the relathionship between working memory and fluid intelligence. *Psichothema, 20,* 780-785.

Burgaleta, M., Johnson, W., Waber, D. P., Colom, R., & Karama, S. (2014). Cognitive ability changes and dynamics of cortical thickness development in healthy children and adolescents. *NeuroImage, 84,* 810-819.

Burgess, G. C., Gray, J. R., Conway, A. R., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General, 140,* 674-692.

Buschkuehl, M., Hernandez-Garcia, L., Jaeggi, S. M., Bernard, J. A., & Jonides, J. (2014). Neural effects of short-term training on working memory. *Cognitive, Affective, & Behavioral Neuroscience, 14,* 147-160.

Buschkuehl, M., & Jaeggi, S. M. (2010). Improving intelligence: A literature review. *Swiss medical weekly, 140,* 266-272.

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience, 2,* S167-S179.

Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences, 108,* 11252-11255.

Byrne, B. M. (1998*). Structural equation modelling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Erlbaum: Mahwah.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6,* 42-57.

Canales-Rodríguez, E. J., Melie-García, L., & Iturria-Medina, Y. (2009). Mathematical description of q-space in spherical coordinates: Exact q-ball imaging. *Magnetic Resonance in Medicine, 61,* 1350-1367.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews, 26,* 321-352.

Caroni, P., Donato, F., & Muller, D. (2012). Structural plasticity upon learning: regulation and functions. *Nature Reviews Neuroscience, 13,* 478-490.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431.

Carroll, J. B. (1993). *Human cognitive abilities.* Cambridge: Cambridge University Press.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of experimental child psychology, 33,* 386-404.

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological psychology, 54,* 241-257.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development?. *Trends in cognitive sciences, 9,* 104-110.

Cattell, R. B. (1940). A culture-free intelligence test. I. *Journal of Educational Psychology, 31*, 161-179.

Cattell, R. B. (1971). *Abilities: their structure, growth, and action.* Boston: Houghton Mifflin.

Cattell, R. B. (1987). *Intelligence: its Structure, Growth and Action: its Structure, Growth and Action*. New York: North-Holland.

Ceccarelli, A., Rocca, M. A., Pagani, E., Falini, A., Comi, G., & Filippi, M. (2009). Cognitive learning is associated with gray matter changes in healthy human individuals: a tensor-based morphometry study. *Neuroimage, 48,* 585-589.

Ceci, S. (1991). How much does schooling influence intellectual development and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703-722.

Chance, S. A., Casanova, M. F., Switala, A. E., & Crow, T. J. (2008). Auditory cortex asymmetry, altered minicolumn spacing and absence of ageing effects in schizophrenia. *Brain, 131,* 3178-3192.

Chen, C. H., Fiecas, M., Gutiérrez, E. D., Panizzon, M. S., Eyler, L. T., Vuoksimaa, E., ... & Kremen, W. S. (2013). Genetic topography of brain morphology. *Proceedings of the National Academy of Sciences, 110,* 17089-17094.

Chen, C. H., Gutierrez, E. D., Thompson, W., Panizzon, M. S., Jernigan, T. L., Eyler, L. T., ... & Dale, A. M. (2012). Hierarchical genetic organization of human cortical surface area. *Science, 335,* 1634-1636.

Chiang, M. C., Barysheva, M., McMahon, K. L., de Zubicaray, G. I., Johnson, K., Montgomery, G. W., ... & Thompson, P. M. (2012). Gene network effects on brain microstructure and intellectual performance identified in 472 twins. *The Journal of Neuroscience, 32,* 8732-8745.

Chiang, M. C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., ... & Thompson, P. M. (2009). Genetics of brain fiber architecture and intellectual performance. *The Journal of Neuroscience, 29,* 2212-2224.

Choi, Y. Y., Shamosh, N. A., Cho, S. H., DeYoung, C. G., Lee, M. J., Lee, J. M., ... & Lee, K. H. (2008). Multiple bases of human intelligence revealed by cortical thickness and neural activation. *The journal of neuroscience, 28,* 10323-10329.

Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence, 40,* 531-542.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112,* 155-159.

Collin, G., & van den Heuvel, M. P. (2013). The Ontogeny of the Human Connectome Development and Dynamic Changes of Brain Connectivity Across the Life Span. *The Neuroscientist, 19,* 616-628.

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography, 18,* 192-205.

Colom, R. (1998). *Psicología de las diferencias individuales. Teoría y Práctica.* Madrid: Pirámide.

Colom, R. (2007). Intelligence? What Intelligence? *Behavioral & Brain Sciences, 30,* 155-56.

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why?. *Intelligence, 36,* 584-606.

Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence, 33,* 623-642.

Colom, R., Burgaleta, M., Román, F. J., Karama, S., Álvarez-Linera, J., Abad, F. J., ... & Haier, R. J. (2013b). Neuroanatomic overlap between intelligence and cognitive factors: Morphometry methods provide support for the key role of the frontal lobes. *NeuroImage, 72*, 143-152.

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the g-factor of intelligence. *Neuroimage, 31,* 1359-1365.

Colom, R., Jung, R. E., & Haier, R. J. (2007). General intelligence and memory span: Evidence for a common neuro-anatomic framework. *Cognitive Neuropsychology, 24,* 867–878.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010b). Human intelligence and brain networks. *Dialogues in clinical neuroscience, 12,* 489-501.

Colom, R., Quiroga, M. Á., Shih, P. C., Martínez, K., Burgaleta, M., Martínez-Molina, A., ... & Ramírez, I. (2010a). Improvement in working memory is not related to increased intelligence scores. *Intelligence, 38,* 497-505.

Colom, R., Quiroga, M. Á., Solana, A. B., Burgaleta, M., Román, F. J., Privado, J., ... & Karama, S. (2012). Structural changes after videogame practice related to a brain network associated with intelligence. *Intelligence, 40,* 479-489.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32,* 277-296.

Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., ... & Jaeggi, S. M. (2013a). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence, 41,* 712-727.

Colom, R., & Thompson, P. M. (2011). Intelligence by Imaging the Brain. In Chamorro-Premuzic, T., von Stumm, S. and Furhan, A (Eds.) *The Wiley-Blackwell handbook of individual differences,* pp. 330-350.

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., ... & Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences, 96,* 10422-10427.

Conway, A. R., & Getz, S. J. (2010). Cognitive ability: Does working memory training enhance intelligence?. *Current Biology, 20,* R362-R364.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review, 12,* 769-786.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences, 7,* 547-552.

Cook, T. K., Appleton, H., Conner, R. F., Shaffer, A., Tamkin, G., & Weber, S. J. (1975). *"Sesame Street" Revisited*. New York: Russell Sage Foundation.

Costa Jr., P. T., & McCrae, R. R. (1992). *The Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory: Professional Manual*. Psychological Assessment Resources, Odessa, Florida.

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

Curlik II, D. M., & Shors, T. J. (2011). Learning increases the survival of newborn neurons provided that learning is difficult to achieve and successful. *Journal of cognitive neuroscience, 23,* 2159-2170.

Dahlin, E., Bäckman, L., Neely, A. S., & Nyberg, L. (2009). Training of the executive component of working memory: Subcortical areas mediate transfer effects. *Restorative Neurology and Neuroscience, 27,* 405–419.

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008a). Transfer of learning after updating training mediated by the striatum. *Science, 320,* 1510-1512.

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. (2008b). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long- term maintenance. *Psychology and Aging, 23,* 720–730.

Davis, S. W., Dennis, N. A., Buchler, N. G., White, L. E., Madden, D. J., & Cabeza, R. (2009). Assessing the effects of age on long white matter tracts using diffusion tensor tractography. *Neuroimage, 46,* 530-541.

de Acedo Lizarraga, M. L. S., Ugarte, M. D., Iriarte, M. D., & de Acedo Baquedano, M. T. S. (2003). Immediate and long-term effects of a cognitive intervention on intelligence, self-regulation, and academic achievement. European journal of psychology of education, 18, 59-74.

Deary, I. J. (2000). Simple information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge: Cambridge University Press.

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63,* 453-482.

Deary, I. J., Austin, E. J., & Caryl, P. G. (2000). Testing versus understanding human intelligence. *Psychology, Public Policy, and Law, 6,* 180-190.

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11,* 201-211.

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: follow-up of the 1932 Scottish Mental Survey. *Intelligence, 28,* 49-55.

Dennis, E. L., Jahanshad, N., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Hickie, I. B., ... & Thompson, P. M. (2013). Development of brain structural connectivity between ages 12 and 30: A 4-Tesla diffusion imaging study in 439 adolescents and adults. *Neuroimage, 64,* 671-684.

Dennis, E. L., & Thompson, P. M. (2013). Mapping connectivity in the developing brain. *International Journal of Developmental Neuroscience, 31,* 525-542.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333,* 959-964.

Dolan, C. V., Colom, R., Abad, F. J., Wicherts, J. M., Hessen, D. J., & van de Sluis, S. (2006). Multi-group covariance and mean structure modeling of the relationship between the WAIS-III common factors and sex and educational attainment in Spain. *Intelligence, 34,* 193-210.

Dostrovsky, J. O., Levy, R., Wu, J. P., Hutchison, W. D., Tasker, R. R., & Lozano, A. M. (2000). Microstimulation-induced inhibition of neuronal firing in human globus pallidus. *Journal of neurophysiology, 84,* 570-574.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature, 427,* 311-312.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Büchel, C., & May, A. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. *The Journal of Neuroscience, 26,* 6314-6317.

Draganski, B., & May, A. (2008). Training-induced structural changes in the adult human brain. *Behavioural brain research, 192,* 137-142.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in neurosciences, 23,* 475-483.

Elliott, R., Sahakian, B. J., Matthews, K., Bannerjea, A., Rimmer, J., & Robbins, T. W. (1997). Effects of methylphenidate on spatial working memory and planning in healthy young adults. *Psychopharmacology, 131,* 196-206.

Engle, R. W., Kane, M. J., Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In Miyake, A. & Shah, P. (Eds.), *Models of working memory* (pp. 102−134). Cambridge: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General, 128,* 309-331.

Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø., Larsen, V. A., & Walhovd, K. B. (2010). Effects of memory training on cortical thickness in the elderly. *Neuroimage, 52,* 1667-1676.

Escorial, S., Román, F. J., Martínez, K., Burgaleta, M., Karama, S., & Colom, R. (2015). Sex differences in neocortical structure and cognitive performance: a surface-based morphometry study. *NeuroImage, 104,* 355-365.

Estrada, E., Ferrer, E., Abad, F. J., Román, F. J., & Colom, R. (2015). A general factor of intelligence fails to account for changes in tests' scores after cognitive practice: A longitudinal multi-group latent-variable study. *Intelligence, 50,* 93-99.

Eysenck, H. J. (1995). Can we study intelligence using the experimental method?. *Intelligence, 20,* 217-228.

Fan, Y., Shi, F., Smith, J. K., Lin, W., Gilmore, J. H., & Shen, D. (2011). Brain anatomical networks in early human brain development. *Neuroimage, 54,* 1862-1871.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes. *Structural Equation Modeling, 6,* 56-83.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39,* 175-191.

Feczko, E., Augustinack, J. C., Fischl, B., & Dickerson, B. C. (2009). An MRI-based method for measuring volume, thickness and surface area of entorhinal, perirhinal, and posterior parahippocampal cortex. *Neurobiology of aging, 30,* 420-431.

Fernández-Ballesteros, R., Botella, J., Zamarrón, M. D., Molina, M. Á., Cabras, E., Schettini, R., & Tárraga, L. (2012). Cognitive plasticity in normal and pathological aging. *Clinical interventions in aging, 7,* 15-25.

Fernández-Ballesteros, R., Zamarrón, M. D., Tárraga, L., Moya, R., & Iñiguez, J. (2003). Learning potential in healthy, mild cognitive impairment subjects and in Alzheimer patients: A research project in Spain. *European Psychologist, 8,* 148-160.

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62,* 774-781.

Fischl, B. & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences, 97,* 11050-11055.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., ... & Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron, 33,* 341-355.

Fleischman, D. A., Leurgans, S., Arfanakis, K., Arvanitakis, Z., Barnes, L. L., Boyle, P. A., ... & Bennett, D. A. (2014). Gray-matter macrostructure in cognitively healthy older persons: associations with age and cognition. *Brain Structure and Function, 219,* 2029-2049.

Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., and Collins, D.L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage, Volume 47, Supplement 1, July 2009, Page S102* Organization for Human Brain Mapping. Annual Meeting.

Forstmann, B., van den Wildenberg, W., & Ridderinkhof, K. (2008). Neural mechanisms, temporal dynamics, and individual differences in interference control. *Journal of Cognitive Neuroscience, 20,* 1854-1865.

Frith, C. D. (2007). The social brain? *Philosophical Transactions of the Royal Society, 362,* 671–678.

Garber, H. L. (1988). *The Milwaukee Project: Preventing Mental Retardation in Children At Risk*. American Association on Mental Retardation, 1719 Kalorama Rd., NW, Washington, DC 20009.

Garcia-Moriyon, F., Rebollo, I., & Colom, R. (2005). Evaluating Philosophy for Children: A meta-analysis. *Thinking: The journal of philosophy for children, 17*, 14-22.

García-Pentón, L., Fernández, A. P., Iturria-Medina, Y., Gillon-Dowens, M., & Carreiras, M. (2014). Anatomical connectivity changes in the bilingual brain. *NeuroImage, 84,* 495-504.

Gardner, H. (2004). Audiences for the theory of multiple intelligences. *The Teachers College Record, 106*, 212-220.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: the role of individual differences in neural plasticity as an explanatory mechanism. *Psychological review, 109,* 116-136.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 23,* 9240-9245.

Gazzaniga, M. S. (Ed.). (2004). *The cognitive neurosciences*. Boston: MIT press.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature neuroscience, 2,* 861-863.

Giedd, J. N., & Rapoport, J. L. (2010). Structural MRI of pediatric brain development: what have we learned and where are we going?. *Neuron, 67,* 728-734.

Giurgen, C. (1973). The 'nootropic' approach to the pharmacology of the integrative activity of the brain. *Conditioned Reflex, 8,* 108−115.

Gläscher, J., Rudrauf, D., Colom, R., Paul, L. K., Tranel, D., Damasio, H., & Adolphs, R. (2010). Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences, 107,* 4705-4709.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America, 101,* 8174-8179.

Gogtay, N., & Thompson, P. M. (2010). Mapping gray matter development: implications for typical development and vulnerability to psychopathology. *Brain and cognition, 72,* 6-15.

Gong, G., Rosa-Neto, P., Carbonell, F., Chen, Z. J., He, Y., & Evans, A. C. (2009). Age-and gender-related differences in the cortical anatomical network. *The Journal of neuroscience, 29,* 15684-15693.

Goto, Y., & Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature neuroscience, 8,* 805-812.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24,* 13-23.

Gottfredson, L. S. (2004). Intelligence: is it the epidemiologists' elusive" fundamental cause" of social class inequalities in health? *Journal of personality and social psychology, 86,* 174-199.

Gray, J., Chabris, C., & Braver, T. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6,* 316–322.

Green, R. L., Hoffman, L. T., Morse, R., Hayes, M. E., & Morgan, R. F. (1964). *The educational status of children in a district without public schools* (Cooperative Research Project No. 2321). Washington, DC: Office of Education, U.S. Department of Health, Education, and Welfare.

Guilford, J. P. (1988). Some changes in the structure-of-intellect model. *Educational and Psychological Measurement, 48,* 1-4.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179-203.

Gustafsson, J. E. (2001). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. *International Education Journal, 2,* 166−186.

Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V., Meuli, R., & Thiran, J. (2006). Imaging the brain neuronal network with diffusion MRI: a way to understand its global architecture. *In ISMRM 14th SCIENTIFIC MEETING, Seattle, Washington, USA, 6-12 May 2006* (Vol. 14, No. EPFL-CONF-91003, p. 436). IEEE.

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., ... & Grant, P. E. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. *Proceedings of the National Academy of Sciences, 107,* 19067-19072.

Haier, R. J. (Ed.). (1993). *Cerebral glucose metabolism and intelligence.* Norwood, NJ: Ablex. ed.

Haier, R. J. (2014). Increased intelligence is a myth (so far). *Frontiers in systems neuroscience, 8.* Article 34.

Haier, R. J., Karama, S., Leyba, L., & Jung, R. E. (2009). MRI assessment of cortical thickness and functional activity changes in adolescent girls following three months of practice on a visual-spatial task. *BMC research notes, 2,* 174.

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., ... & Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence, 12,* 199-217.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11,* 236–242.

Hampson, M., Driesen, N. R., Skudlarski, P., Gore, J. C., & Constable, R. T. (2006). Brain connectivity related to working memory performance. *The Journal of neuroscience, 26,* 13338-13343.

Hashemi, R. H., Bradley, W. G., Lisanti, C. (2004). *MRI: The Basics, 2nd edition*. Philadelphia: Williams & Wilkins.

Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., & Bauman, A. (2007). Physical activity and public health—Updated recommendation for adults from the American college of sports medicine and the American heart association. *Circulation, 116,* 1081–1093.

Hativa, N. (1988). Computer-based drill and practice in arithmetic: Widening the gap between high-and low-achieving students. *American Educational Research Journal, 25,* 366-397.

Heinzel, S., Schulte, S., Onken, J., Duong, Q. L., Riemer, T. G., Heinz, A., ... & Rapp, M. A. (2014). Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Aging, Neuropsychology, and Cognition, 21,* 146-173.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Herrnstein, R. J., Nickerson, R. S., de Sanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist,* 41, 1279-1289.

Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., & Van Essen, D. (2010). Similar patterns of cortical expansion during human development and evolution. *Proceedings of the National Academy of Sciences, 107,* 13135-13140.

Hindin, S. B., & Zelinski, E. M. (2012). Extended practice and aerobic exercise interventions benefit untrained cognitive outcomes in older adults: A meta-analysis. *Journal of the American Geriatrics Society, 60,* 136–141.

Hof, P. R., & Morrison, J. H. (2004). The aging brain: morphomolecular senescence of cortical circuits. *Trends in neurosciences, 27,* 607-613.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental science, 12,* F9-F15.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Woldman (Ed.), *Handbook of intelligence: Theories, measurement, and applications*. New York: Wiley, pp. 267-300.

Hornung, C., Brunner, M., Reuter, R. A. P., & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence, 39,* 210–221.

Hsu, N. S., Buschkuehl, M., Jonides, J., & Jaeggi, S. M. (2013, November). *Potential Mechanisms Underlying Working Memory Training and Transfer.* Poster presented at the Psychonomic Society Annual Meeting, Toronto, Ontario.

Hu, L. T. & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling, 6,* 1-55.

Hunt, E. (2011). *Human intelligence*. Cambridge University Press.

Hunt, E. (2012a). *Improving Intelligence*. Manuscript not published.

Hunt, E. (2012b). What makes nations intelligent?. *Perspectives on Psychological Science, 7,* 284-306.

Hunt, E., & Jaeggi, S. M. (2013). Challenges for research on intelligence. *Journal of Intelligence, 1,* 36-54.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of comparative Neurology, 387,* 167-178.

Ilg, R., Wohlschläger, A. M., Gaser, C., Liebau, Y., Dauner, R., Wöller, A., ... & Mühlau, M. (2008). Gray matter increase induced by practice correlates with task-specific activation: a combined functional and morphometric magnetic resonance imaging study. *The Journal of Neuroscience, 28,* 4210-4215.

Irwing, P., Hamza, A., Khaleefa, O., & Lynn, R. (2008) Effects of abacus training on the intelligence of Sudanese children. *Personality and Individual Differences, 45,* 694-696.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences, 96,* 9379-9384.

Ishibashi, T., Dakin, K. A., Stevens, B., Lee, P. R., Kozlov, S. V., Stewart, C. L., & Fields, R. D. (2006). Astrocytes promote myelination in response to electrical impulses. *Neuron, 49,* 823-832.

Iturria-Medina, Y. (2013). Anatomical brain networks on the prediction of abnormal brain states. *Brain connectivity, 3,* 1-21.

Iturria-Medina, Y., Canales-Rodriguez, E. J., Melie-Garcia, L., Valdes-Hernandez, P. A., Martinez-Montes, E., Alemán-Gómez, Y., & Sánchez-Bornot, J. M. (2007). Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *NeuroImage, 36,* 645-660.

Iturria-Medina, Y., Sotero, R. C., Canales-Rodríguez, E. J., Alemán-Gómez, Y., & Melie-García, L. (2008). Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. *Neuroimage, 40,* 1064-1076.

Izawa, M.R., French, M.D., Hedge, A. (2011) Shining new light on the Hawthorne illumination experiments. *Human Factors, 53,* 528-547.

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., ... & Weiner, M. W. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging, 27,* 685-691.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105,* 6829-6833.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short-and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences, 108,* 10081-10086.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010b). The concurrent validity of the N-back task as a working memory measure. *Memory, 18,* 394-412.

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & cognition, 42,* 464-480.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010a). The relationship between n-back performance and matrix reasoning: Implications for training and transfer. *Intelligence, 38,* 625–635.

James, W. (1980). *The Principles of Psychology.* Boston, MA: Holt.

Jaušovec, N., & Jaušovec, K. (2012). Working memory training: improving intelligence–changing brain activity. *Brain and Cognition, 79,* 96-106.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jensen, A. R. (2006). *Clocking the mind.* London: Elsevier.

Jensen, A. R. (1989). Raising IQ without increasing g? A review of "The Milwaukee Project: Preventing Mental Retardation in Children at Risk." *Developmental Review, 9,* 234-258.

Jeurissen, B. (2012). *Improved analysis of brain connectivity using high angular resolution diffusion MRI.* (Doctoral dissertation, Universiteit Antwerpen, Belgium).

Johnson, W. (2013). Whither intelligence research? *Journal of Intelligence, 1,* 25-35.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33,* 393-416.

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence, 32,* 95-107.

Jones, D. K., Simmons, A., Williams, S. C., & Horsfield, M. A. (1999). Non-invasive assessment of axonal fiber connectivity in the human brain via diffusion tensor MRI. *Magnetic Resonance in Medicine, 42,* 37-41.

Jonides, J., Schumacher, E., Smith, E., Lauber, E., Awh, E., Minoshima, S., & Koeppe, R. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Cognitive Neuroscience, Journal of, 9,* 462-475.

Jöreskog, K. (1993). Testing structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 294−315). Newbury Park: Sage.

Joshi, A. A., Lepore, N., Joshi, S. H., Lee, A. D., Barysheva, M., Stein, J. L., ... & Thompson, P. M. (2011). The contribution of genes to cortical thickness and volume. *Neuroreport, 22,* 101-105.

Juan-Espinosa, M. (1997). *Geografía de la inteligencia humana: las aptitudes cognitivas.* Madrid: Pirámide.

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behavioral and Brain Sciences, 30,* 135-154.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 615–622.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189-217.

Karama, S., Ad-Dab'bagh, Y., Haier, R. J., Deary, I., Lyttelton, … & the Brain Development Cooperative Group (2009). Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds. *Intelligence. 37,* 145-155.

Karama, S., Colom, R., Johnson, W., Deary, I. J., Haier, R., Waber, D. P., ... & Brain Development Cooperative Group. (2011). Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6 to 18. *Neuroimage, 55,* 1443-1453.

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental science, 12,* 978-990.

Karnath, H. O. (2001). New insights into the functions of the superior temporal cortex. *Nature Reviews Neuroscience, 2,* 568-576.

Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., & Shah, P. (2014). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in human neuroscience, 8.* Article 242.

Keizer, A. W., Verschoor, M., Verment, R. S., & Hommel, B. (2010). The effect of gamma enhancing neurofeedback on the control of feature bindings and intelligence measures. *International Journal of Psychophysiology, 75,* 25-32.

Kelley, A. E., Smith-Roe, S. L., & Holahan, M. R. (1997). Response-reinforcement learning is dependent on N-methyl-D-aspartate receptor activation in the nucleus accumbens core. *Proceedings of the National Academy of Sciences, 94,* 12174-12179.

Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., ... & Evans, A. C. (2005). Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage, 27,* 210-221.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *Neuroreport, 8,* 3581-3585.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352-358.

Kirwan, C. B., & Stark, C. E. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory, 14,* 625-633.

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of educational research, 78,* 85-123.

Klein, A., & Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in neuroscience, 6.* Art. 171.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in cognitive sciences, 14,* 317-324.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of clinical and experimental neuropsychology, 24,* 781-791.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences, USA, 103,* 10155–10162.

Kompier, M. A. (2006). The "Hawthorne effect" is a myth, but what keeps the story going? *Scandinavian Journal of Work, Environment and Health, 32,* 402–412.

Kriegeskorte, N., Lindquist, M. A., Nichols, T. E., Poldrack, R. A., & Vul, E. (2010). Everything you never wanted to know about circular analysis, but were afraid to ask. *Journal of Cerebral Blood Flow & Metabolism, 30,* 1551-1557.

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37,* 347-364.

Kühn, S., Gleich, T., Lorenz, R. C., Lindenberger, U., & Gallinat, J. (2014). Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. *Molecular psychiatry, 19,* 265-271.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162–181.

Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *The Journal of Neuroscience, 33,* 8705-8715.

Kwok, V., Niu, Z., Kay, P., Zhou, K., Mo, L., Jin, Z., ... & Tan, L. H. (2011). Learning new color names produces rapid increase in gray matter in the intact adult human cortex. *Proceedings of the National Academy of Sciences, 108,* 6686-6688.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence, 14,* 389-433.

Lange, N., Froimowitz, M. P., Bigler, E. D., Lainhart, J. E., & Brain Development Cooperative Group. (2010). Associations between IQ, total and regional brain volumes, and demography in a large normative sample of healthy children and adolescents. *Developmental neuropsychology, 35,* 296-317.

Lerch, J. P., & Evans, A. C. (2005). Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage, 24,* 163-173.

Li, S. C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15,* 155-163.

Li, S.C., Schmiedek, F.,Huxhold, O., Rocke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging, 23,* 731–742.

Lipton, M. L. (2008). *Totally accessible MRI. A users guide to principles, technology, and applications*. New York: Springer.

Locurto, C. (1991). Beyond IQ in preschool programs?. *Intelligence, 15,* 295-312.

Lövdén, M., Bodammer, N. C., Kühn, S., Kaufmann, J., Schütze, H., Tempelmann, C., ... & Lindenberger, U. (2010). Experience-dependent plasticity of white-matter microstructure extends into old age. *Neuropsychologia, 48,* 3878-3883.

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., & Bäckman, L. (2013). Structural brain plasticity in adult learning and development. *Neuroscience & Biobehavioral Reviews, 37,* 2296-2310.

Luchtman, D. W., & Song, C. (2013). Cognitive enhancement by omega-3 fatty acids from child-hood to old age: findings from animal and clinical studies. *Neuropharmacology, 64,* 550-565.

Lynch, G., Palmer, L. C., & Gall, C. M. (2011). The likelihood of cognitive enhancement. *Pharmacology Biochemistry and Behavior, 99,* 116-129.

Lyttelton, O., Boucher, M., Robbins, S., & Evans, A. (2007). An unbiased iterative group registration template for cortical surface analysis. *NeuroImage, 34,* 1535-1544.

Lyttelton, O. C., Karama, S., Ad-Dab'bagh, Y., Zatorre, R. J., Carbonell, F., Worsley, K., & Evans, A. C. (2009). Positional and surface area asymmetry of the human cerebral cortex. *NeuroImage, 46,* 895-903.

MacDonald, D., Kabani, N., Avis, D., & Evans, A. C. (2000). Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage, 12,* 340-356.

Mackey, A. P., Whitaker, K. J., & Bunge, S. A. (2012). Experience-dependent plasticity in white matter microstructure: reasoning training alters structural connectivity. *Frontiers in neuroanatomy, 6,* 32.

Mackworth, J. F. (1959). Paced memorizing in a continuous task. *Journal of Experimental Psychology, 58,* 206-211.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences, 97,* 4398-4403.

Martínez, K. (2014). *Where in the brain is intelligence? An exhaustive analysis across gray matter properties, neuroimaging protocols and matched samples*. (Doctoral dissertation, Universidad Autónoma de Madrid).

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to 'simple'short-term storage?. *Intelligence, 39,* 473-480.

Martínez, K., Madsen, S., Joshi, A. A., Shantanu, J., Román, F. J., … & Colom, R. (2015). Reproducibility of Brain-Cognition Relationships Using Three Cortical Surface Based Protocols: An Exhaustive Analysis Based On Cortical Thickness. *Human brain mapping, 36,* 3227–3245.

Martínez, K., Solana, A. B., Burgaleta, M., Hernández-Tamames, J. A., Álvarez-Linera, J., Román, F. J., ... & Colom, R. (2013). Changes in resting-state functionally connected parietofrontal networks after videogame practice. *Human brain mapping, 34,* 3143-3157.

Matthews, G., Roberts, R. D., & Zeidner, M. (2004). Seven myths about emotional intelligence. *Psychological Inquiry, 15,* 179-196.

May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in cognitive sciences, 15,* 475-482.

May, A., Hajak, G., Gänssbauer, S., Steffens, T., Langguth, B., Kleinjung, T., & Eichhammer, P. (2007). Structural brain alterations following 5 days of intervention: dynamic aspects of neuroplasticity. *Cerebral Cortex, 17,* 205-210.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 817-835.

McGaugh, J.L. & Roozendaal, B. (2009). Drug enhancement of memory consolidation: Historical perspective and neurobiological implications. *Psychopharmacology, 202,* 3–14.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1-10.

McKey, R. H. (1985). *The Impact of Head Start on Children, Families, and Communities*. A. N. Smith and S. S. Aitken (eds.). Washington, D.C.: U.S. Department of Health and Human Services.

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., & Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science, 323,* 800-802.

Mechelli, A., Price, C. J., Friston, K. J., & Ashburner, J. (2005). Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging Reviews, 1,* 105-113.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49,* 270-291.

Messick, S., & Jungeblut, A. (1981). Time and method in coaching for the SAT. *Psychological Bulletin, 89,* 191-216.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 9−100.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640.

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence, 37,* 327–328.

Mori, S., Crain, B. J., Chacko, V. P., & Van Zijl, P. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of neurology, 45,* 265-269.

Mori, S., & Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron, 51,* 527-539.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review, 18,* 46–60.

Muthén, L.K. and Muthén, B.O. (1998-2008). *Mplus (Version 5.1)*. Los Angeles, CA: Muthén & Muthén.

Narr, K. L., Woods, R. P., Thompson, P. M., Szeszko, P., Robinson, D., Dimtcheva, T., ... & Bilder, R. M. (2007). Relationships between IQ and regional cortical gray matter thickness in healthy adults. *Cerebral Cortex, 17,* 2163-2171.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... & Urbina, S. (1996). Intelligence: knowns and unknowns. *American psychologist, 51,* 77-101.

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews, 33,* 1004-1023.

Nikolaidis, A., Voss, M. W., Lee, H., Vo, L. T., & Kramer, A. F. (2014). Parietal plasticity after training with a complex video game is associated with individual differences in improvements in an untrained working memory task. Frontiers in human neuroscience, 8. Art. 169.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American psychologist, 67,* 130-159.

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language, 51,* 80-96.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31,* 167-193.

Oelhafen, S., Nikolaidis, A., Padovani, T., Blaser, D., Koenig, T., & Perrig, W. J. (2013). Increased parietal activity after training of interference control. *Neuropsychologia, 51,* 2781-2790.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9,* 97-113.

Østby, Y., Tamnes, C. K., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2009). Heterogeneity in subcortical brain development: a structural magnetic resonance imaging study of brain maturation from 8 to 30 years. *The Journal of neuroscience, 29,* 11772-11782.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping, 25,* 46-59.

Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., ... & Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex,* bhp026.

Parker, G. J., & Alexander, D. C. (2003, January). Probabilistic Monte Carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. In *Information Processing in Medical Imaging* (pp. 684-695). Springer Berlin Heidelberg.

Parker, G. J., Haroon, H. A., & Wheeler-Kingshott, C. A. (2003). A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *Journal of Magnetic Resonance Imaging, 18,* 242-254.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review Neuroscience, 28,* 377-401.

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in cognitive sciences, 9,* 60-68.

Paus, T. (2010). Growth of white matter in the adolescent brain: myelin or axon?. *Brain and cognition, 72,* 26-35.

Perrig, W. J., Hollenstein, M., & Oelhafen, S. (2009). Can we improve fluid intelligence with training on working memory in persons with intellectual disabilities? *Journal of Cognitive Education and Psychology, 8,* 148–164.

Plomin. R, de Fries J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral Genetics* (Worth, New York), 5th Ed.

Pol, H. E. H., Schnack, H. G., Posthuma, D., Mandl, R. C., Baaré, W. F., van Oel, C., ... & Kahn, R. S. (2006). Genetic contributions to human brain morphology and intelligence. *The Journal of Neuroscience, 26,* 10235-10242.

Pollitt, E., Gorman, K. S., Engie, E L., Martorell, R., & Rivera, J. (1993). Early supplementary feeding and cognition. *Monographs of the Society for Research in Child Development, 58* (Serial No. 235).

Posner, M. I., & Rothbart, M. K. (2007). *Educating the Human Brain.* Washington, DC: American Psychological Association.

Posthuma, D., De Geus, E. J., Baare, W. F., Hulshoff Pol, H. E., Kahn, R. S., & Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience. 5,* 83–84.

Prabhakaran, V., Rypma, B., Narayanan, N. S., Meier, T. B., Austin, B. P., Nair, V. A., ... & Gabrieli, J. D. (2011). Capacity-speed relationships in prefrontal cortex. *PloS one, 6,* e27504.

Quiroga, M. Á., Escorial, S., Román, F. J., Morillo, D., Jarabo, A., Privado, J., ... & Colom, R. (2015). Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can!. *Intelligence, 53,* 1-7.

Quiroga, M. Á., Herranz, M., Gómez-Abad, M., Kebir, M., Ruiz, J., & Colom, R. (2009). *Video-games: Do they require general intelligence?.* Computers & Education, 53, 414-418.

Quiroga, M. Á, Román, F. J., Catalán, A., Rodríguez, H., Ruiz, J., Herranz, M., ... & Colom, R. (2011). Videogame Performance (Not Always) Requires Intelligence. *International Journal of Online Pedagogy and Course Design (IJOPCD), 1,* 18-32.

Rae, C., Digney, A. L., McEwan, S. R., & Bates, T. C. (2003). Oral creatine monohydrate supplementation improves brain performance: a double–blind, placebo–controlled, cross–over trial. *Proceedings of the Royal Society of London B: Biological Sciences, 270,* 2147-2150.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science, 241,* 170-176.

Ramey, C. T. (1992). High-risk children and IQ: altering intergenerational patterns. *Intelligence, 16,* 239-256.

Ramey, C. T., MacPhee, D., & Yeates, K. O. (1982). Preventing develnprnental retardation: A general systems model. In *How and How Much Can Intelligence Be Increased*. D. K. Detterman and R. J. Sternberg (eds.). Norwood, New Jersey: Ablex Publishing Corp., pp. 67-1 19.

Ramon y Cajal, S. (1904). *Textura del Sistema Nervioso del Hombre y de los Vertebrados (Vol. 2)*. Madrid.

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S., Ellis, C., Shakeshaft, C., ... & Price, C. J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature, 479,* 113-116.

Raven, J. C. (1988). *Standard progressive matrices.* Weinheim: Beltz.

Raven, J., Raven, J. C., & Court, J. H. (2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Pearson Assessment: San Antonio, TX.

Raznahan, A., Shaw, P., Lalonde, F., Stockman, M., Wallace, G. L., Greenstein, D., ... & Giedd, J. N. (2011). How does your cortex grow?. *The Journal of Neuroscience, 31,* 7174-7177.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142,* 359.

Reeve, C. L., & Lam, H. (2005). The psychometric paradox of practice effects due to retesting: Measurement invariance and stable ability estimates in the face of observed score changes. *Intelligence, 33,* 535–549.

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015, March 16). Is Education Associated With Improvements in General Cognitive Ability, or in Specific Skills?. *Developmental Psychology*. Advance online publication.

Rivkin, M. J. (2000). Developmental neuroimaging of children using magnetic resonance techniques. *Mental retardation and developmental disabilities research reviews, 6,* 68-80.

Román, F. J., Abad, F. J., Escorial, S., Burgaleta, M., Martínez, K., Álvarez-Linera, J., ... & Colom, R. (2014). Reversed hierarchy in the brain for general and specific cognitive abilities: A morphometric analysis. *Human brain mapping, 35,* 3805-3818.

Román, F. J., García-Rubio, M. J., Privado, J., Kessel, D., López-Martín, S., Martínez, K., ... & Colom, R. (2015). Adaptive working memory training reveals a negligible effect of emotional stimuli over cognitive processing. *Personality and Individual Differences, 74,* 165-170.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., ... & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: a coordinate-based meta-analysis. *Neuroimage, 60,* 830-846.

Rudebeck, S. R., Bor, D., Ormond, A., O'Reilly, J. X., & Lee, A. C. (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PloS one, 7,* e50431.

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the national Academy of Sciences of the United States of America, 102,* 14931-14936.

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews, 39,* 34-50.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist, 55,* 68-78.

Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., & Assaf, Y. (2012). Learning in the fast lane: new insights into neuroplasticity. *Neuron, 73,* 1195-1203.

Sahakian, B., & Morein-Zamir, S. (2007). Professor's little helper. *Nature, 450,* 1157−1159.

Salgado-Pineda, P., Baeza, I., Pérez-Gómez, M., Vendrell, P., Junqué, C., Bargalló, N., & Bernardo, M. (2003). Sustained attention impairment correlates to gray matter decreases in first episode neuroleptic-naive schizophrenic patients. *Neuroimage, 19,* 365-375.

Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working memory training on executive functioning. *Frontiers in Human Neuroscience, 6.* Article 166.

Sanabria-Diaz, G., Melie-García, L., Iturria-Medina, Y., Alemán-Gómez, Y., Hernández-González, G., Valdés-Urrutia, L., ... & Valdés-Sosa, P. (2010). Surface area and cortical thickness descriptors reveal different attributes of the structural human brain networks. *Neuroimage, 50,* 1497-1510.

Satorra, A. & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In Alexander von Eye and Clifford C. Clogg (Eds.), *Latent variables analysis: Applications to developmental research*. (pp. 399-419). Thousands Oaks, CA: Sage Publications Inc.

Schmidt-Wilcke, T., Rosengarth, K., Luerding, R., Bogdahn, U., & Greenlee, M. W. (2010). Distinct patterns of functional and structural neuroplasticity associated with learning Morse code. *Neuroimage, 51,* 1234-1241.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of personality and social psychology, 86,* 162-173.

Schmidt, H., Jogia, J., Fast, K., Christodoulou, T., Haldane, M., Kumari, V., & Frangou, S. (2009). No gender differences in brain activation during the N-back task: An fMRI study in healthy individuals. *Human brain mapping, 30*, 3609-3615.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Lindenberger, U., & Wilhelm, O. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1089 –1096.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in aging neuroscience, 2*-27.

Schmithorst, V. J., Wilke, M., Dardzinski, B. J., & Holland, S. K. (2005). Cognitive functions correlate with white matter architecture in a normal pediatric population: a diffusion tensor MRI study. *Human brain mapping, 26*, 139-147.

Schnack, H. G., van Haren, N. E., Brouwer, R. M., Evans, A., Durston, S., Boomsma, D. I., ... & Pol, H. E. H. (2014). Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cerebral Cortex, bht357*.

Schoenthaler, S. J., Amos, S. P., Eysenck, H. J., Peritz, E., & Yudkin, J. (1991). Controlled trial of vitamin-mineral supplementation: Effects on intelligence and performance. *Personality and Individual Differences, 12*, 351-362.

Schoenthaler, S. J., Bier, I. D., Young, K., Nichols, D., & Jansenns, S. (2000). The effect of vitamin-mineral supplementation on the intelligence of American schoolchildren: a randomized, double-blind placebo-controlled trial. *The Journal of Alternative and Complementary Medicine, 6*, 19-29.

Schoenthaler, S. J., Doraz, W. E., & Wakefield, J. A. (1986). The testing of various hypotheses as explanations for the gains in National Standardized Academic Test Scores in the 1978–1983 New York City Nutrition Policy Modification Project. *International Journal of Biosocial Research, 8*, 196-203.

Scholz, J., Klein, M. C., Behrens, T. E., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature neuroscience, 12*, 1370-1371.

Schwaighofer, M., Fischer, F., & Bühner, M. (2015) Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators. *Educational Psychologist, 50*, 138-166.

Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: increasing cognitive and affective executive control through emotional working memory training. *PLoS One, 6*, e24372.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., ... & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature, 440*, 676-679.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., ... & Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *The Journal of Neuroscience, 28*, 3586-3594.

Sherbondy, A. J., Dougherty, R. F., Ananthanarayanan, R., Modha, D. S., & Wandell, B. A. (2009). Think global, act local; projectome estimation with BlueMatter. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009* (pp. 861-868). Springer Berlin Heidelberg.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective?. *Psychological bulletin, 138*, 628-654.

Shors, T. J. (2014). The adult brain makes new neurons, and effortful learning keeps them alive. *Current Directions in Psychological Science, 23*, 311-318.

Sled, J.G., Zijdenbos, A.P., and Evans, A.C. (1998). A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data. *IEEE Transactions on Medical Imaging 17,* pp. 87–97.

Smith, M. E., & Farah, M. J. (2011). Are prescription stimulants "smart pills"? The epidemiology and cognitive neuroscience of prescription stimulant use by normal healthy individuals. *Psychological bulletin, 137,* 717-741.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping, 17,* 143-155.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., ... & Behrens, T. E. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage, 31,* 1487-1505.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23,* S208-S219.

Solfrizzi, V., Panza, F., Frisardi, V., Seripa, D., Logroscino, G., … & Pilotto, A. (2011) Diet and Alzheimer's disease risk factors or prevention: The current evidence. *Expert Review of Neurotherapeutics, 11,* 677-708.

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature neuroscience, 6,* 309-315.

Spearman, C. (1904). " General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15,* 201-292.

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics, 2,* 145-162.

Stanford Center on Longevity (2009). Expert Consensus on Brain Health. [website]; 2009 [updated 2009; cited 2009 Jul 6]; Available from: http://longevity.stanford.edu/mymind/cognitiveagingstatement.

Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence, 41,* 341-357.

Sternberg, R. J. (1988). *The triarchic mind.* London: Penguin Books.

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences, 105,* 6791-6792.

Stough, C., Camfield, D., Kure, C., Tarasuik, J., Downey, L., Lloyd, J., ... & Reynolds, J. (2011). Improving general intelligence with a nutrient-based pharmacological intervention. *Intelligence, 39,* 100-107.

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences, 53,* 44-49.

Sur, M., & Rubenstein, J. L. (2005). Patterning and plasticity of the cerebral cortex. *Science, 310,* 805-810.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261-288.

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., ... & Kawashima, R. (2010). Training of working memory impacts structural connectivity. *The Journal of neuroscience, 30,* 3297-3303.

Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Fukushima, A., & Kawashima, R. (2011). Working memory training using mental calculation impacts regional gray matter of the frontal and parietal regions. *PLoS One, 6,* e23175.

Tamnes, C. K., Fjell, A. M., Østby, Y., Westlye, L. T., Due-Tønnessen, P., Bjørnerud, A., & Walhovd, K. B. (2011). The brain dynamics of intellectual development: Waxing and waning white and gray matter. *Neuropsychologia, 49,* 3605-3611.

Tamnes, C. K., Østby, Y., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2010). Brain maturation in adolescence and young adulthood: regional age-related changes in cortical thickness and white matter volume and microstructure. *Cerebral Cortex, 20,* 534-548.

Tang, C. Y., Eaves, E. L., Ng, J. C., Carpenter, D. M., Mai, X., Schroeder, D. H., ... & Haier, R. J. (2010). Brain networks for working memory and factors of intelligence assessed in males and females with fMRI and DTI. *Intelligence, 38,* 293-303.

Tang, Y. Y., & Posner, M. I. (2009). Attention training and attention state training. *Trends in cognitive sciences, 13,* 222-227.

Tapia, J. A., & de la Red Fadrique, I. (2007). Evaluar "para" el aprendizaje, aprender para estar motivado: el orden de los factores sí afecta al product [*Assessment for learning, to learn to be motivated: the order of factors alters the product*]. *REOP-Revista Española de Orientación y Psicopedagogía, 18,* 241-253.

Taubert, M., Draganski, B., Anwander, A., Müller, K., Horstmann, A., Villringer, A., & Ragert, P. (2010). Dynamic properties of human brain structure: learning-related changes in cortical areas and associated fiber connections. *The Journal of Neuroscience, 30,* 11670-11677.

Taya, F., Sun, Y., Babiloni, F., Thakor, N., & Bezerianos, A. (2015). Brain enhancement through cognitive training: a new insight from brain connectome. *Frontiers in systems neuroscience, 9,* 44.

te Nijenhuis, J., van Vianen, A. E.M., & van der Flier, H. (2007). Score gains on g loaded tests: No g. *Intelligence, 35,* 283–300.

Thomas, A. G., Marrett, S., Saad, Z. S., Ruff, D. A., Martin, A., & Bandettini, P. A. (2009). Functional but not structural changes associated with learning: an exploration of longitudinal voxel-based morphometry (VBM). *Neuroimage, 48,* 117-125.

Thomas, C., & Baker, C. I. (2013a). Teaching an adult brain new tricks: a critical review of evidence for training-dependent structural plasticity in humans. *NeuroImage, 73,* 225-236.

Thomas, C., & Baker, C. I. (2013b). On evidence, biases and confounding factors: Response to commentaries. *NeuroImage, 73,* 265-267.

Thompson, P. M., Hayashi, K. M., Sowell, E. R., Gogtay, N., Giedd, J. N., Rapoport, J. L., ... & Toga, A. W. (2004). Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. *Neuroimage, 23,* S2-S18.

Thompson, T. W., Waskom, M. L., Garel, K. L., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., . . . Gabrieli, J. D. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS One, 8,* e63614.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental science, 12,* 106-113.

Thurstone, L. (1938). *Primary mental abilities*. Psychometric Monographs.

Thurstone, L. L., & Thurstone, Th. G. (1968). *Aptitudes mentales primarias (PMA) manual.* Madrid, Spain: Tea Ediciones, S.A.

Tidwell, J. W., Dougherty, M. R., Chrabaszcz, J. R., Thomas, R. P., & Mendoza, J. L. (2014). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychonomic bulletin & review, 21,* 620-628.

Toga, A. W., & Thompson, P. M. (2005). Genetics of brain structure and intelligence. *Annual Review Neuroscience, 28,* 1-23.

Tohka, J., Zijdenbos, A., & Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage, 23,* 84-97.

Tournier, J., Calamante, F., & Connelly, A. (2013). Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR in Biomedicine, 26,* 1775-1786.

Tsuchida, A., & Fellows, L. K. (2009). Lesion evidence that two distinct regions within prefrontal cortex are critical for n-back performance in humans. *Journal of Cognitive Neuroscience, 21,* 2263-2275.

Tuch, D. S. (2002b). *Diffusion MRI of complex tissue structure* (Doctoral dissertation, Harvard Medical School).

Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine, 52,* 1358-1372.

Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., & Wedeen, V. J. (2002a). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine, 48,* 577-582.

Tymofiyeva, O., Hess, C. P., Xu, D., & Barkovich, A. J. (2014). Structural MRI connectome in development: challenges of the changing brain. *The British journal of radiology, 87,* 20140086.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent variable analysis of the relationship between processing and storage. *Memory, 17,* 635–654.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological bulletin, 139,* 352-402.

Valdés-Conroy, B., Sebastián, M., Hinojosa, J. A., Román, F. J., & Santaniello, G. (2014). A close look into the near/far space division: A real-distance ERP study. *Neuropsychologia, 59,* 27-34.

Vallejo-Nágera, A., & Colom, R. (2004). *Tu inteligencia: cómo entenderla y mejorarla.* Madrid: Aguilar.

van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *The Journal of neuroscience, 31,* 15775-15786.

van den Heuvel, M. P., & Sporns, O. (2013). Network hubs in the human brain. *Trends in cognitive sciences, 17,* 683-696.

Vartanian, O., Jobidon, M. E., Bouak, F., Nakashima, A., Smith, I., … & Cheunga, B. (2013). Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. *Neuroscience, 236,* 186–194.

Vendetti, M. S., & Bunge, S. A. (2014). Evolutionary and Developmental Changes in the Lateral Frontoparietal Network: A Little Goes a Long Way for Higher-Level Cognition. *Neuron, 84,* 906-917.

Verhaeghen, P., & Basak, C. (2005). Ageing and switching of the focus of attention in working memory: Results from a modified N-back task. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 58,* 134-154.

Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: How to expand the focus of serial attention from one to four ítems in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1322-1337.

Verstraete, E., Veldink, J. H., Mandl, R. C., Van den Berg, L. H., & van den Heuvel, M. P. (2011). Impaired structural motor connectome in amyotrophic lateral sclerosis. *PloS one, 6,* e24239.

von Bastian, C. C., & Eschen, A. (2015). Does working memory training have to be adaptive?. *Psychological research,* 1-14.

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language, 69,* 36-58.

Vuoksimaa, E., Panizzon, M. S., Chen, C. H., Fiecas, M., Eyler, L. T., Fennema-Notestine, C., ... & Kremen, W. S. (2014). The genetic association between neocortical volume and general cognitive ability is driven by global surface area rather than thickness. *Cerebral Cortex,* bhu018.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory. *Cognitive, Affective, & Behavioral Neuroscience, 3,* 255-274.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-WISC-IV*. San Antonio, Texas: Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV)*. San Antonio, Texas: NCS Pearson.

Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y. I., Dai, G., ... & de Crespigny, A. J. (2008). *Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers.* Neuroimage, 41, 1267-1277.

Whalley, L. J., & Deary, I. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal, 322* (7 April), 1-5.

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *Neuroimage, 87*, 120-126.

Wilhelm, O., & Engle, R. W. (2005). *Handbook of understanding and measuring intelligence*. California: Sage.

Willis, S. L., & Schaie, K. W. (2009). Cognitive training and plasticity: theoretical perspective and methodological consequences. *Restorative neurology and neuroscience, 27,* 375-389.

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., ... & Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage, 53,* 1135-1146.

Woollett, K., & Maguire, E. A. (2011). Acquiring "the Knowledge" of London's layout drives structural brain changes. *Current biology, 21*, 2109-2114.

World Medical Association (2008). Declaration of Helsinki — Ethical principles for medical research involving human subjects. *59th WMA General Assembly, Seoul,* Korea.

Yang, J. J., Yoon, U., Yun, H. J., Im, K., Choi, Y. Y., Lee, K. H., ... & Lee, J. M. (2013). Prediction for human intelligence using morphometric characteristics of cortical surface: partial least square analysis. *Neuroscience, 246,* 351-361.

Yap, P. T., Fan, Y., Chen, Y., Gilmore, J. H., Lin, W., & Shen, D. (2011). Development trends of white matter connectivity in the first years of life. *PloS one, 6,* e24678.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature methods, 8,* 665-670.

Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences, 108,* 8873-8878.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in neurosciences, 34,* 515-525.

Yeo, B. T., Krienen, F. M., Eickhoff, S. B., Yaakub, S. N., Fox, P. T., Buckner, R. L., ... & Chee, M. W. (2014). Functional specialization and flexibility in human association cortex. *Cerebral Cortex, bhu217.*

Yu, C., Li, J., Liu, Y., Qin, W., Li, Y., Shu, N., ... & Li, K. (2008). White matter tract integrity and intelligence in patients with mental retardation and healthy adults. *Neuroimage, 40,* 1533-1541.

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010b). Network-based statistic: identifying differences in brain networks. *NeuroImage, 53,* 1197-1207.

Zalesky, A., Fornito, A., Harding, I. H., Cocchi, L., Yücel, M., Pantelis, C., & Bullmore, E. T. (2010a). Whole-brain anatomical networks: does the choice of nodes matter?. *Neuroimage, 50,* 970-983.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience, 15,* 528-536.

Zhang, J., Wang, J., Wu, Q., Kuang, W., Huang, X., He, Y., & Gong, Q. (2011). Disrupted brain connectivity networks in drug-naive, first-episode major depressive disorder. *Biological psychiatry, 70,* 334-342.

Zhao, L., Boucher, M., Rosa-Neto, P., & Evans, A. C. (2013). Impact of scale space search on age-and gender-related changes in MRI-based cortical morphometry. *Human brain mapping, 34,* 2113-2128.

Zhou, D., Lebel, C., Treit, S., Evans, A., & Beaulieu, C. (2015). Accelerated longitudinal cortical thinning in adolescence. *NeuroImage, 104,* 138-145.

Zigler, E., & Muenchow, S. (1992). *Head Start: The Inside Story of America's Most Successful Educational Experiment.* New York: Basic Books.

Zou, Q., Ross, T. J., Gu, H., Geng, X., Zuo, X. N., Hong, L. E., ... & Yang, Y. (2013). Intrinsic resting-state activity predicts working memory brain activation and behavioral performance. *Human brain mapping, 34,* 3204-3215.